REPUBLIC OF TÜRKİYE
ALTINBAŞ UNIVERSITY
Institute of Graduate Studies
Information Technologies

NETWORK TRAFFIC CLASSIFICATION USING MACHINE LEARNING OR DEEP LEARNING OR DATA MINING

Enas Saad Jawad AL-NIDAWI

Master's Thesis

Supervisor

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Istanbul, 2023

NETWORK TRAFFIC CLASSIFICATION USING MACHINE LEARNING OR DEEP LEARNING OR DATA MINING

Enas Saad Jawad AL-NIDAWI

Information Technologies

Master's Thesis

ALTINBAŞ UNIVERSITY

2023

The thesis titled NETWORK TRAFFIC CLASSIFICATION USING MACHINE LEARNING OR DEEP LEARNING OR DATA MINING. Prepared by "ENAS SAAD JAWAD AL-NIDAWI" and submitted on 5/04/2023 has been **accepted unanimously** for the degree of Master of Science in Information Technology

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Supervisor

Thesis Defence Committee Members:

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering Altinbaş University	_____
Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN	Department of Computer Engineering, Altinbaş University	_____
Asst. Prof. Dr. Zeynep ALTAN	Department of Engineering and Architecture Beykent University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to the Graduate Education Institute: ___/___/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Enas Saad Jawad AL-NIDAWI

Signature

DEDICATION

I would like to thank my supervisor Asst. Prof. Dr. Ayça KURNAZ TÜRKBEN , and specıal thanks to my family specıally my mother who gave me the encouragement I needed throughout thıs poroess and to my husband and my chıldren to support me all the tıme.

ABSTRACT

NETWORK TRAFFIC CLASSIFICATION USING MACHINE LEARNING OR DEEP LEARNING OR DATA MINING

AL-NIDAWI, Enas Saad Jawad

M.Sc. Information Technologies, Altınbaş University,

Supervisor: Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Date: April /2023

Pages: 63

Malware is the general name of software that is coded to infiltrate and damage computers without the knowledge of computer users. It is placed to provide unauthorized access to information networks and to be used for different purposes against the will of its users. In terms of countries where malware is found, Far East countries are more dense in terms of quantity, while this rate is relatively less in European countries. We can say that the reason for this is the establishment of the legal infrastructure and the implementation of prevention studies more intensively. It is possible to say that the rates of malware infecting computers are higher in Turkey and Taiwan, especially in China.

In this study, new method applied to classify the data in computer network to the classes normal and abnormal. The proposed method based genetic algorithm which are used to train the SVM and presented 99.64% accuracy.

Keywords: CNN, GA, SVM, Computer Network Traffic Classification, Abnormal.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

CNN : Convolutional Neural Network

SVM : Support Vector Machine

NN : Neural Network

ANN : Artificial Neural Network

KNN : K-Nearest Neighbour

1. INTRODUCTION

In today's world, the security and protection of information resources has become more important than ever before. The emergence of new threats in the globalizing and digitalizing world shows itself mostly in the cyber field. Digital services ranging from personal data to infrastructures have become the target of malware, malicious hackers or organized government units. Health data, education data, infrastructure facilities, etc. Many factors have to be constantly protected by countries and used in sound systems [1,2]. In the cyber world, situations such as attacks, terrorism, espionage, crime or bullying are constantly increasing, leaving states, individuals and institutions constantly on alert. Internet of Things, Industry 4.0 etc. The boundaries of the world connected with concepts are becoming increasingly unclear and new areas of struggle are emerging. Cyber security can collapse with the vulnerability of an actor, or an attack on an actor can leave other actors vulnerable. The security of the social, political, economic and cultural structures of the countries has changed into a virtual structure in the 21st century [3,4]. The security of communication networks is extremely important not only in terms of privacy, but also for the flow of life, the continuity of economic life and the functioning of political systems. Events such as the collapse of communication lines during a natural disaster can cause social turmoil and political depression. In this respect, both security and continuity are a direct reflection of cyber security. Computer technologies used in all areas of life facilitate and automate human work and, most importantly, put it into a sustainable structure. It is a reflection of cyber security that all applications used are reliable, stable and accessible at any time. Especially in Turkey, the security and accessibility of e-government applications have gained more importance. The importance of cyber security is increasing day by day, from large systems where sensitive data of citizens are kept to all structures where personal data of employees are stored [5]. Reliability of information has become an area that has increased its importance especially in the last ten years. False or manipulative information spread by social media can cause social unrest, damage political systems and manipulate communities. In this respect, taking precautions, preventing the spread of false information and protecting the social order have come to the fore in recent years as a reflection of cyber security. The field of cyber security is dispersed and covers many different areas. In this respect, it is complex and very difficult to manage. This field, which is too important to be left in the hands of a single

person or expert, should have a center, sub-units affiliated to the center should be established and the authorities, duties and responsibilities of each unit should be determined. It is seen that this field, which was mostly delivered to information processing units in our country, has gained importance and has become independent units. Despite this, it is expected that the development will accelerate and systematic structures will be established strongly. As seen in many incidents, no matter how reliable a system is designed, there is always the possibility of being attacked or damaged. In such cases, the important thing is that the post-attack recovery is fast, the information is recovered reliably and the continuity of the work is maintained without interruption. On the basis of cyber security, it is dependent on the end user. The slightest mistake of the end user can easily undermine a well-structured system, or it is stated that a system, no matter how automatic, is extremely sensitive and vulnerable to the human factor. In this respect, it has become very important for end users to be educated about cyber attacks, to know the types of cyber attacks, and to be sensitive about possible risk factors. As will be examined in detail in the study, special attention was paid to the training of the end user in the strategy and action plan of each country and many activities were created. It can be said that special attention has been given to end-user trainings, especially in government bodies [6].

2. OVERVIEW

2.1 ARTIFICIAL INTELLIGENCE

Building machines that can think has been among the dreams of scientists since the early history of science. In the 1800s, English mathematician and mechanical engineer Charles BOBBGE found some errors in the logarithm tables used at that time. He stated that these erroneous charts would cause problems in engineering calculations and in preparing ship routes in the military and civilian area, which was of great importance at that time. He first came up with the idea that these calculations could be done faster and more advanced with an automatic machine in 1812, and he started the project in 1823. “The machine he called the difference machine would calculate the values of the polynomial functions from the first given value. The project was interrupted because the conditions of that time were not at the level to produce sensitive parts and it had a high budget. Later, although Babbage designed another machine, which he called the Difference Machine-2, it was interrupted for the same reasons. While designing the Babbage Difference Machine, he designed another machine he called the "Analytical Machine". The most important features that distinguish the Analytical Machine from the Difference Machine were its wider purpose and programmable with punch cards. The information exchange between the machine and the user was to be provided by punch cards. Analytical Machine, defined as the first general purpose computer in terms of its structure, could not be completed due to the material and technological impossibilities of the period [7].

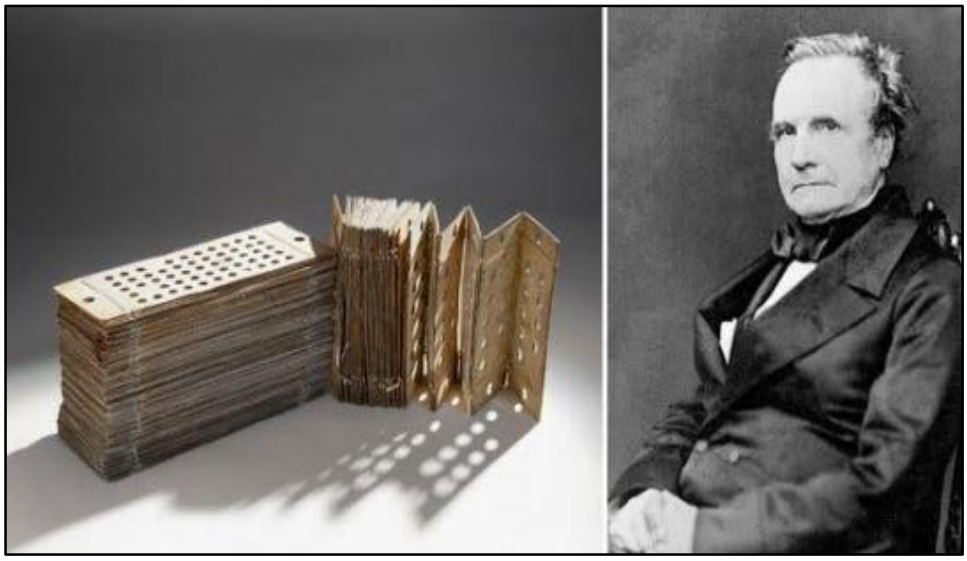

Figure 2.1: AI Father.

English mathematician known for his work on the "Analytical Machine", known as the first general-purpose computer, and Ada LOVELACE, known as the first programmer in history, said in a speech in 1843 that this machine was not built to create something, it was developed to help people if they did not do things that can already be done. According to Lovelace, such a machine, if properly programmed, would be able to compose complex musical works, produce graphics and solve complex mathematical problems, etc. could be used for In the letters he sent to Babbage, Lovelace explained how the machine in question could calculate Bernoulli numbers using a plan consisting of a certain and finite number of steps. 8 This expression is known as the first “computer program” that could be applied to a concrete machine in the history of computers [8].

When artificial intelligence emerged, it fulfilled its duties in the short term. These are complex mathematical calculations where the human brain and computing machines of the era were slow. But in Artificial intelligence, which has risen with the developing technology and utopian goals such as flying cars or talking maid robots, it has been revealed that the main problem is not mathematical calculations, but more intuitive issues such as speech recognition or object finding, which are very easy for a healthy person. Over time, this problem has been addressed and the effort to get closer to the human brain has started from the calculator. First, problems with certain rules and in a certain order were started to be solved. The Deep Tought chess machine, developed at Carnegie Mellon University in the

eighties, became the first machine to do so by beating the chess master in a legal tournament. Later, the Deep Taought team moved to the IBM Research Center and developed Deep Taught 2, also known as Deep Blue. Deep Blue defeated chess world champion Gary Gasparov in 1997 with a score of 3.5 to 2.5 in a six-game game [9]. The idea of thinking machines, which started in the 1800s, has only recently reached the level of realizing heuristic problems. This was because there were an infinite number of versions of routines encountered in everyday life, and the programmer had to program these representations as input to the machine. Thus, theoretically, the machine could reason using rules of logical inference. This approach, known as the Knowladge base (also known as the expert system), which was used especially in the 80s, was insufficient and could not achieve the expected success. An example of projects using this approach is an inference engine written by Lenat and Guha using the CyCL language created with the Cyc Cyc database in 1989 [10]. In his 1933 article "Computable Numbers", English mathematician Alan Turing talked about not only a mathematical system, but also virtual machines that could think and do the necessary calculations instead of humans when the necessary rules were introduced. The Turkish equivalent of the Calculator machine mentioned in the article is "computer counting". It is designed to read the symbols written on a tape, compare them with the symbols entered before, and give a result accordingly and write on the tape in the same way. Making logical transformations shows its similarity to human intelligence. Alan Turing, who published his article "Computer Mechanism and Intelligence" in 1950, came up with the idea of an experiment that would determine whether a machine considered the Turing test is intelligent. Accordingly, if the machine could convince the human questioner that he was human after the conversation he had with this person, this machine (computer) would be considered as intelligent as the human. In the 8-week workshop held at Dartmouth College in New Hampshire in 1956, AI research was discussed and topics such as computers, natural language processing, neural networks, computation theory, abstraction and creativity were covered. The term artificial intelligence was first coined here. An excerpt from the proposal for a conference is as follows: “We propose a 2month, 10-person artificial intelligence study to be conducted at Dartmouth College in Hanover, New Hampshire, in the summer of 1956. The work will proceed on the assumption that every aspect of learning or any other aspect of intelligence can in principle be described so precisely that the machine can simulate it. We will try to find out how machines will use language, how to create abstractions and

concepts, how to solve the types of problems reserved for humans, and how to improve themselves. "We think that if a carefully selected group of scientists work together over the course of a summer, significant progress can be made on one or more of these problems." [11].

Eight approaches to AI are discussed in two dimensions in their work titled “Artificial Intelligence: A Modern Approach” by Stuart Russell and Peter Norvig [12]. These approaches are as follows 10 • Humane behavior: Turing test put forward by Alan Turing can be given as an example. In order to be suitable for this approach, it is expected to know the subjects of natural language processing (NLP), knowledge representation, machine learning, automated reasoning, computer vision and robotics. Obviously, this approach includes the main disciplines of modern artificial intelligence. It is an approach based on machines behaving like humans. • Human thinking: Also known as the cognitive modeling approach. It is based on the ability of machines to think and make decisions like humans and solve problems. It is based on examining the correlation of the outputs received with respect to the inputs given to the machine with the outputs obtained when the same inputs are given to the human. An example of this approach is GPS (General Problem Solver) by Newell and Simon. The similarity of Artificial Intelligence with the human cognitive system and the fact that it is based on this approach reveals that the two issues are constantly intertwined with each other.

i- Rational thinking: It is the approach of transferring mental characteristics of humans to machines by using computer-based computational models. It is based on logic. He says that all problems based on logical basis can be solved with logical patterns, if there is no logical solution, the program will continue in an endless loop. It is also known as the laws of thought approach.

ii- Acting rationally: Acting rationally means acting to achieve one's goals, given one's beliefs or understanding of the world. An agent is a system that senses an environment and acts in that environment. An intelligent agent is someone who acts rationally and logically for his purposes. For example, an agent designed to play games must make moves that increase his chances of winning the game. In creating an intelligent agent, that is, the system, the agent shifts towards designing the best possible decision model under the conditions in which the agent is acting, rather than designing the best possible decision-making model

theoretically. But making the best theoretically possible decision and achieving what is called "perfect rationality" is often not possible in a real setting.

In line with the historical developments and approaches described above, artificial intelligence can be explained as the acquisition of rational behaviors such as interpreting events, benefiting from experiences, making generalizations and learning, which are unique to humans, apart from mathematical operations where the human brain is partially slow. The aim of artificial intelligence is to produce machines that have their own intelligence and can reach results faster for situations where human intelligence is required. In other words, artificial intelligence is a model of the human brain. Artificial intelligence is divided according to the techniques used while modeling. Some of these are shown below.

2.2 EXPERT SYSTEMS

Expert systems are machines that have the knowledge of people who can be qualified as experts in their fields. These systems, which are knowledge-based, help their users in making decisions and reaching results, as experts in the subject, without human support. Expert systems consist of three components. These are the user interface, the inference engine, and the knowledge base. The user interface transmits the questions it receives from the user, who establishes the relationship between the non-expert user and the system, to the inference engine in an appropriate format. It then sends the results obtained from the inference engine to the user as an output. The inference engine works as the brain of the system and works with forward and backward chaining methods. The knowledge base acts as a repository. He keeps the information he receives from other experts on the same subject here. Expert systems are used especially in various subspecialties in medicine and education. Abdülkerim Öncü's doctoral study titled "Development of a Web-Based Instructional Evaluation System with an Expert System Approach" designed a system consisting of modules called ODM (Learning Evaluation System). These modules are respectively teaching module, exam, psychology, multiple criteria, data collection, analysis and reporting modules. Expert Systems were used in the analysis module. The study aimed to increase the efficiency in education by the fact that teaching can be used by both students and teachers without the need for any software knowledge [13].

2.3 COMPUTER VISION

Computer vision is the subject of artificial intelligence to extract meaningful information

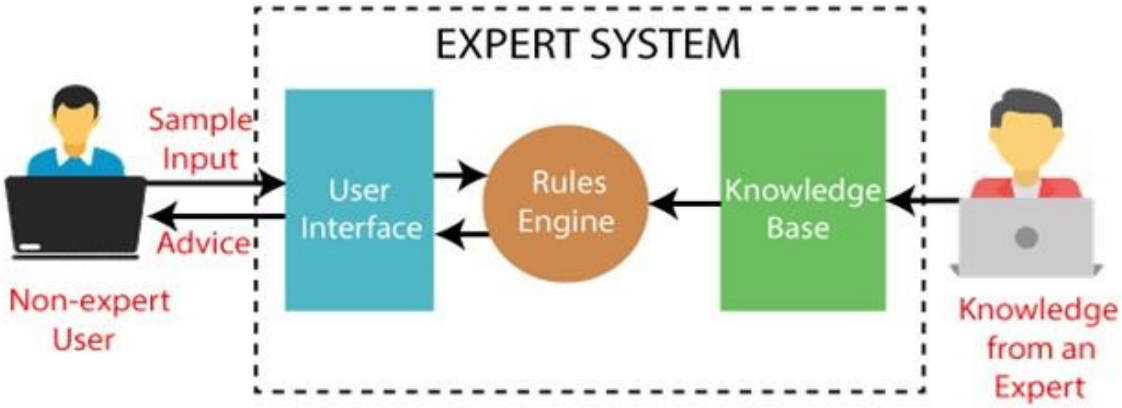

Figure 2.2: Expert System.

from digital data such as photos, pictures, and videos. Computer vision, which deals with everything human vision can do, such as object recognition, understanding the distance between objects, or distinguishing between objects, works better the more data it is trained on. Computer vision, which needs a lot of data for this, performs at an awareness and speed far above human capacity when trained with enough data. For example, the data set of cats and dogs in the Keras library in Python consists of more than 3 million image data. Deep learning is the most used method of computer vision, as it has the opportunity to make hierarchical inferences between layers and to be trained over and over again with a lot of data. The most widely used deep learning algorithm in this regard is convolutional neural networks (CNN). ESAs, consisting of convolution and deposition pairs, process the image pixel by pixel. Computer vision has become one of the most popular areas of artificial intelligence due to the billions of digital images that appear almost everyday thanks to the internet and the advancement of hardware developments that can process these images, and the success rate of the models created has approached 100 [14].

2.4 SPEECH RECOGNITION

Speech recognition is a machine's ability to convert the human voice into a written format. It plays an important role in the lives of individuals with physical disabilities due to its ability to work without requiring a physical (manual) intervention. Unlike voice recognition, it does not focus on just one person's voice, but processes the data it receives from all users. It

converts speech, which is initially in the form of a sound wave, into an electrical signal into a format that can be processed by the hardware itself. It then outputs according to the commands it receives from the user. This is where artificial intelligence comes into play. In time, it can learn the commands it receives from the user and label who the speaker belongs to, generalize the desired data and give similar results, complete the wrong or incomplete input data, or filter the results that it deems harmful. Today, speech recognition-based artificial intelligence systems are used in systems that can receive voice commands in vehicles, in virtual notebooks that enable the user to convert their speech into written format and save time, or in systems that solve the problem without the need for customer service by talking to the customer who has a problem with the purchased product. Alexa, developed by Amazon and released in 2014, and Siri, introduced by Apple in 2011, are the most popular artificial intelligence systems using speech recognition. The most used speech recognition algorithms are Natural Language Processing, Dynamic Time Warping, and Neuron Networks. Dynamic time warping is a method that compares the similarities between time-dependent data and previously known words and finds the best result within certain rules. Neuron networks developed by referring to the neuron network structure in the human brain are used especially in deep learning and play a role in the behaviour of machines such as deriving new information from the information in their hands. Natural language processing is a widely used artificial intelligence method. Understanding the natural language used by people according to its roots and suffixes, separating it and deriving results, it performs the learning process [15].

2.5 MACHINE LEARNING

Machine learning makes it possible to perform tasks that would be insufficient to solve programs written by users. In other words, it is called machine learning to do it by self-learning without waiting for the user to receive it or coding it openly in the program. Learning and duty are not the same thing. Learning is the system acquiring enough skills to perform the task successfully. Machine learning has broken new ground in programming. In classical programming logic, when the rules are introduced to the system and the data is given as input, it produces results suitable for these inputs, while in machine learning, it should be able to learn the rules that enable it to produce appropriate results for different inputs by introducing the inputs and the expected outputs of these inputs to the system. It

also finds these rules as statistically based. With the developing hardware and increasing data numbers over the years, classical machine learning algorithms have been insufficient. As a result, the demand for deep learning, which is less math and statistics-based and more programming-based, has increased.

It is a branch of information technologies and software systems science, which is also known as smart, has features such as benefiting from the coding language, understanding, evaluating, and solving problems, that is, learning to make decisions by using software. The more comprehensive definition of artificial intelligence is; It is a science that makes tools equipped with features that will enable them to perform tasks that require intelligence such as getting information, grasping, seeing, understanding and making decisions. Machine Learning, on the other hand, is an existing AI application based on the idea that machines should be able to access data and let them learn for themselves. The current model in which Machine Learning happens is known as Neural Network. A Neural Network, a computer program designed to work by classifying information the same way the human brain does, can make decisions or make predictions. The addition of a feedback loop makes learning possible. They are increasingly used in real estate analysis and forecasting, finding correlations beyond what is possible with traditional regression models. Natural Language Processing is another AI application that uses Machine Learning to help machines understand the nuances of human language and give answers in a way that anyone with intermediate intelligence can understand. This is especially important for use in chatbots and customer service information technology, and home assistants such as Alexa, which have entered the residential market in the last few years, are also starting to become popular. A third area of AI development is computer detection/recognition systems from digital images or video, European and US privacy concerns such as the facial recognition scandals at Sidewalk Laboratories in Toronto, government intervention on ethical grounds at Kings Cross in London, and the US ban of Huawei. This technology is heavily used in China, although it is restricted by laws, data security and fears of data manipulation, and it cannot be denied that this technology can have a big impact on real estate. The development of facial recognition technology will provide an understanding of the use of space and the areas will be adjusted according to personal preferences. It will also enable retailers to better understand the customer in the store or mall and customize their shopping experience accordingly [16].

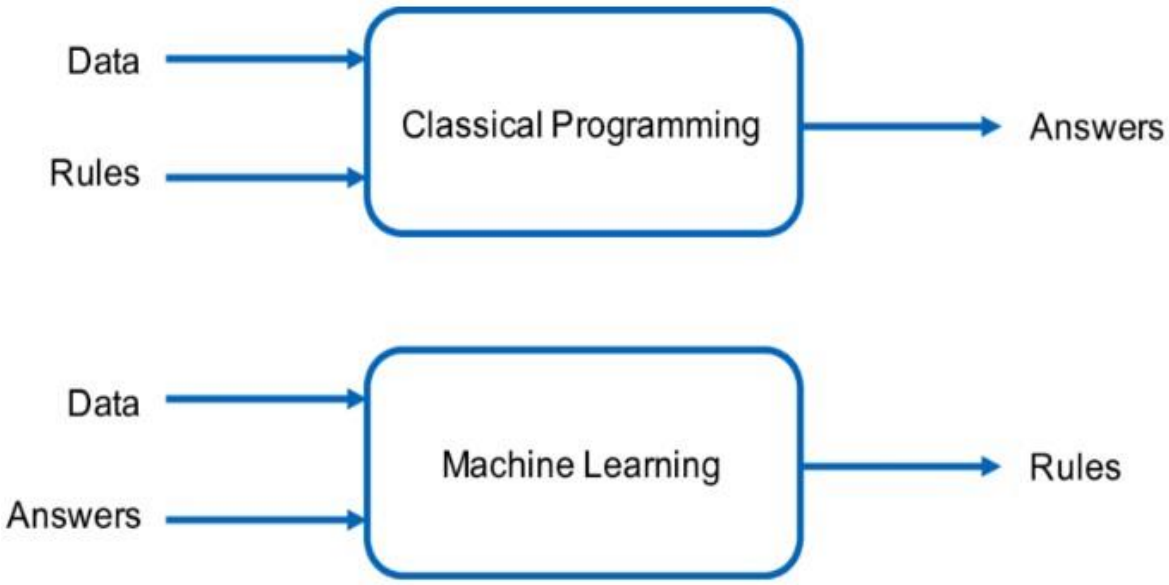

Figure 2.3: Machine Learning.

2.6 TYPES OF MACHINE LEARNING

Before moving on to Deep Learning, it is useful to explain machine learning in detail. There are many machine learning systems available. They can be divided according to whether the information given as input is a belled or unlabelled, or how much of it is a belled, whether all the input data is introduced to the system in parts or gradually, or it can be divided into sample-based and model-based. These criteria can be used not completely separately from each other, but also in a hybrid way [17].

2.6.1 Supervised Learning

If the data feeding the Supervised Learning Model is a belled, that is, it contains the expected results, it is called supervised learning. The simplest supervised learning tasks are classification and relevance. The classification task indicates which class it belongs to when the new data is entered into the trained model. The simplest example is that the mailbox can separate mail into essential and junk. The other task context gives a numerical output. By looking at the given inputs and labels, it makes a numerical prediction for the data requested from the model. For example, it is a commitment task for a model trained with a data set where the attributes of the houses are varied (tax amount, crime rate of the location, age of

the house, etc.) and the label value determines the rent, to predict the rent of a house with certain attributes. The correlation problem can also be used for classification. The ability to give the percentage value of which class the desired data belongs to is called logistic relevance. The most known supervised learning algorithms; Nearest neighbors, Linear coupling, Logistic coupling, Support vector machines (SVM), Decision trees and Random forests, ANN and deep learning are explained in detail in section [18].

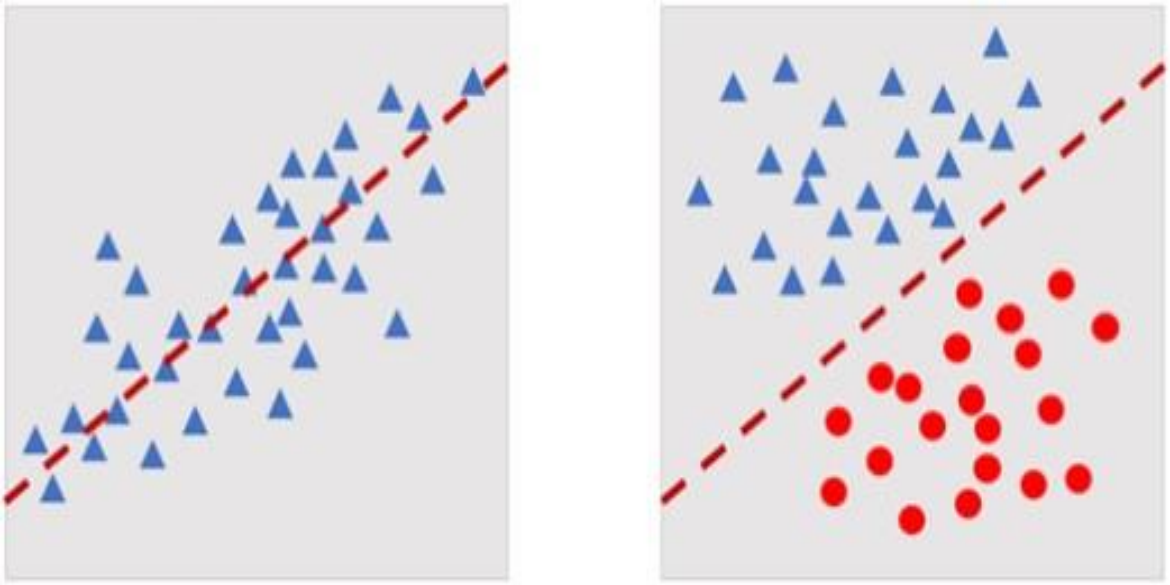

Figure 2.4: SVM for Classification.

a. K-Nearest Neighbors Method K-Nearest Neighbor algorithm is a supervised learning algorithm. It is used in both classification and regression problems. It is an algorithm that is used to determine the class of the data whose class is unknown by comparing the distances of the data to the other data in the training set, according to which it is more similar. The logic in KNN is based on similar data32 being close to each other. While normal classification algorithms assign class information over a certain class, the KNN algorithm creates a classifier for each data point over the nearest neighbor data to that point and classifies the values. KNN is also called Lazy Student Algorithm because of iterative classifiers [19]. The KNN algorithm is an algorithm

that has high accuracy and is resistant to noise in the data, although the applied procedures are long. Some distance functions used. The KNN algorithm works as follows:

i. Data is loaded.

ii. The k parameter, that is, the number of neighbours to be examined, is determined.

iii. For each of the determined neighbours, the distance is calculated by running the determined distance function. Calculated offsets are added to an array.

iv. The obtained distance array is ordered from smallest to largest, and the first element of the array is taken.

v. Label of the selected k neighbours is selected to the new data.

The number of neighbours selected in the KNN algorithm is important. For this, the algorithm is run several times for different k values, and the model that can make the most appropriate estimation for the new data is tried to be selected. The KNN method is a machine learning algorithm that is easy to implement. The K-Nearest Neighbors method is widely preferred in classification problems due to its advantages such as being easy to train the model, easy to calculate based on mathematical formulas, and being sensitive to mixed data sets with noise. Despite these advantages of the KNN algorithm, there are also some disadvantages. The distances calculated for each point significantly increase the processing load, time and cost. Some of these disadvantages are that the operation needs to be tried several times to find the optimal value due to the random number of k neighbors selected. 33 The KNN method is a type of machine learning that is frequently used in object and pattern identification. In the classification study with the KNN algorithm made by ÇINAR et al., the model trained from the IONOLABTEC data was carried out to determine the damages caused by ,99 Marmara earthquakes and 99 solar eclipses [20].

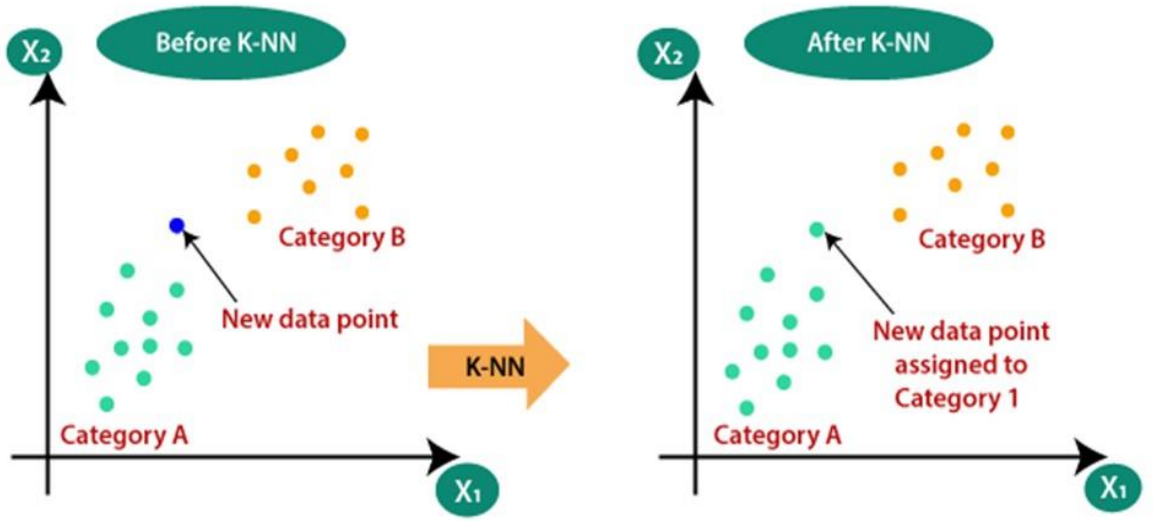

Figure 2.5: KNN.

b. Decision Trees: The decision tree method is a type of machine learning that is frequently used in classification and regression problems in machine learning and also forms the basis of the Random Forests method. It clearly represents the data and the model in flowcharts and, as the name suggests, looks like a tree from root to branch. In the tree diagram, branches give the rules and leaves give the output values. The tree structure starting from the root branches until the model outputs. The leaves represent the output of the model. Each branch structure and root represent an attribute that can be used in classification. Decision trees are known as transparent box models because of their ease of use and clarity. It is quite clear on what basis or according to which rules the classification of the model is made. Like other data trees, they work from top to bottom. It works with a simple "if else" loop structure for binary classifiers. The decision tree algorithm recursively divides the data set into groups and performs the learning process from these groups. The grouping process continues until the node value in each group reaches the target tag value. Decision trees can work with continuous and categorical data. While creating the tree flowchart, it is of great importance to determine the properties to be assigned to the root value and sub-branches. There are two important methods developed to determine this. These are the information gain and the Gini index. Information weight is the measure of whether the information in the groups obtained after clustering the data carries enough information about the classes. A high information weight means that the uncertainty in the data is reduced. Root and branch values are determined according to the amount of information in the subsets and the entropy, that is,

the disorder value. The model tries to maximize the amount of information gain. The formula below shows the information gain [21].

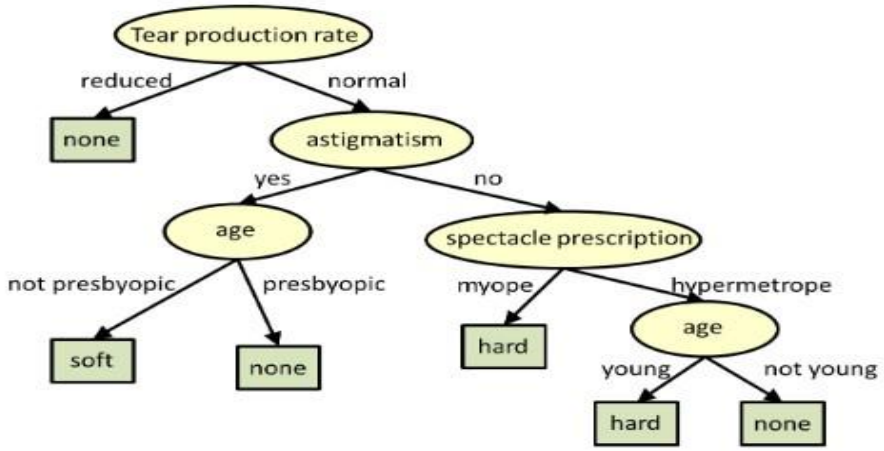

Figure 2.6: DT.

c. Ensemble Learning and Random Forests the concept of ensemble learning is the most different type of machine learning in terms of its structure and use. Regardless of classification or connection, it is a type of machine learning where the model to be created is not a single model, but the results created by more than one model within itself, by voting by the main model. Since the main model consists of the information of more than one model, its performance is higher and more reliable than a single model. The submodels used can be of different types. In the model created by different types of classifiers, different algorithms will work, and the errors made by each estimator will be different, so the generalization ability of the model will be high. The main model created with ensemble learning is more flexible and more useful than single models developed with other types of machine learning. That's why the Ensemble learning method is one of the most used and popular branches of machine learning. The ensemble learning method consists of two separate methods, depending on whether the models used are used simultaneously or sequentially. These are the packing method (bagging) and the boosting method (boosting). In the packaging method, each submodel that makes up the main model works

simultaneously and is trained with a subset of the data set. In the increment method, the sub-models work sequentially and take the information learned by the previous models as input. The Random Forests method is an ensemble learning method that can work with any of the packaging and augmentation methods. It is also a type of supervised machine learning. It is called Random Forests because all estimators in the ensemble learning method consist of decision trees. The performance of the parallel model increases with the increase in the number of decision trees in the main model. It consists of a set of decision trees trained with subsets of the dataset. It can be used for classification and coupling problems. The dataset on which each decision tree runs is such that it is a random subset of its main data. The predictions generated by each decision tree trained simultaneously are collected and voted on by the main model. The prediction with the most votes emerges as the prediction of the main model. This is called a hard vote classifier. For example, to buy a new computer, assume a Random Forests model that outputs the user to buy or not. Let it be assumed that the model consists of 4 decision trees and processes RAM, weight, price and brand attributes. At the end of the training, the model can be voted according to the output of each decision tree and with the majority vote, it can give a buy or buy output with a higher success than a single decision tree. In regression problems, the average of more than one continuous prediction formed by the sub-models is given as the output of the main model. 36 Random forests method is more successful than Decision Trees method because it consists of many decision trees instead of a single decision tree. Problems such as over fitting in decision trees have been solved in random forests. The method, which is used in many fields, especially in the field of health and finance, has made new product suggestions to its customers to buy by taking advantage of the past behaviour of its consumers in the field of e-commerce in recent years [22].

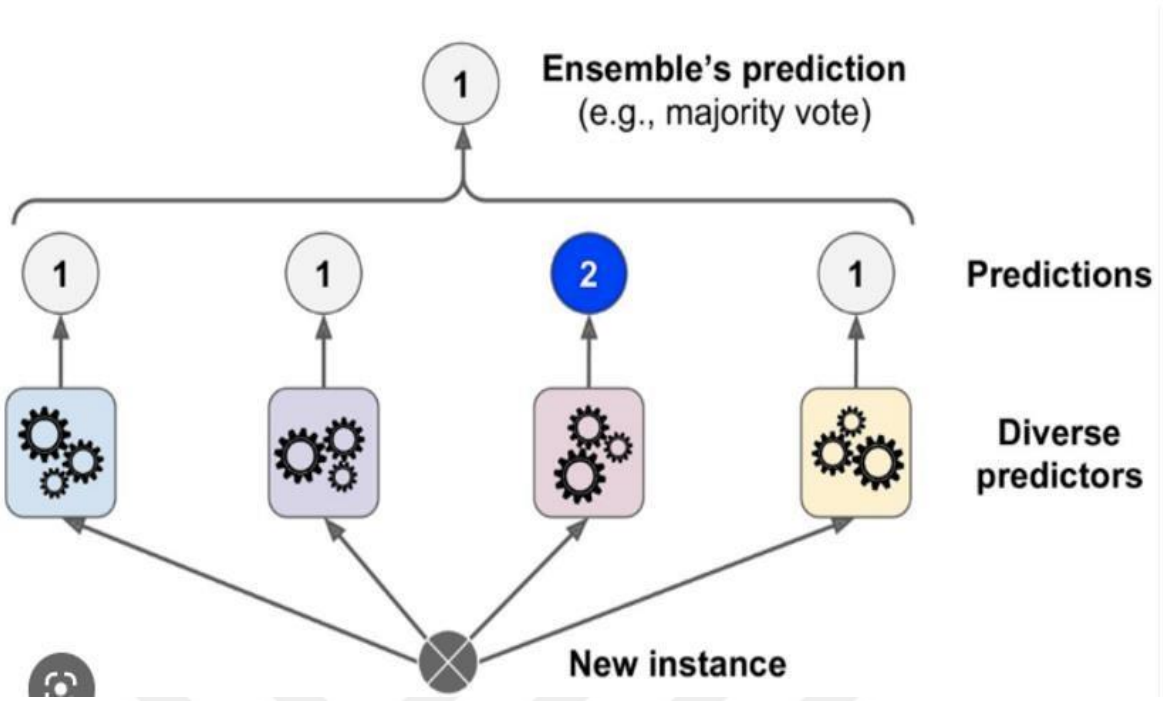

Figure 2.7: Ensemble Learning.

d. Support Vector Machines Support Vector Machines are one of the most used and popular types of machine learning. It is frequently used in linear or non-linear classification problems and regression. In linearly separable classification problems, support vector machines try to separate the classes from each other as much as possible, while drawing the boundary for separating the classes called decision boundary or hyperplane. The purpose of this is to increase the success rate of the model in new data. Data points close to the determined decision boundary are called support vectors. An N-dimensional plane (n-1) has hyperplanes. Support vector machines (DVMs) are frequently used in binary or multiple classifications [23].

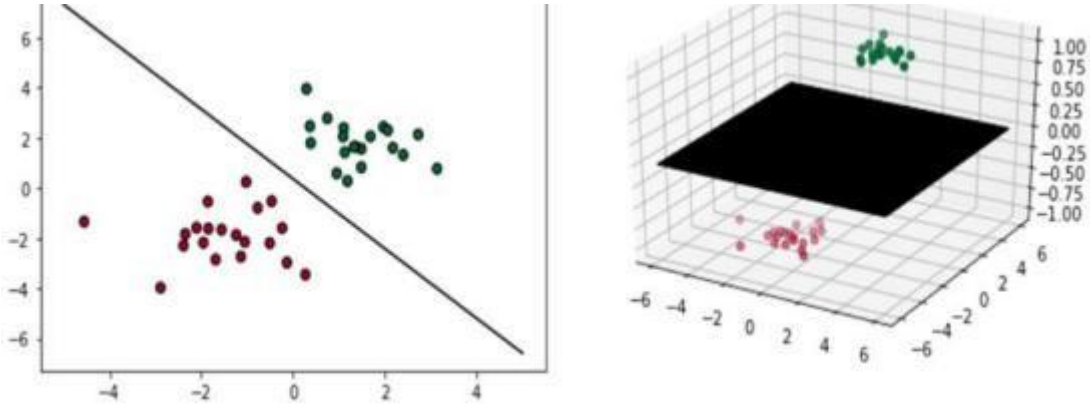

Figure 2.8: SVM Process 1.

While classifying, SVM tries to find the hyperplane that will maximize the margin value, that is, the distance between the data. However, data points may not always be next to the designated street, but sometimes above it or even within the margin. If the margin value determination is made taking into account this internal data, it is called hard constraint. The ability of the model to generalize in hard constraint is not developed. At such times, it is possible to be more flexible in determining the hyperplane. The decision limit can be drawn by ignoring the values that are not in large numbers on or inside the street. This indicates that the DVM performs well against cluttered and noisy data. This is called soft constraint. The chart below clearly shows the hard and soft constraints. The hyper parameter "C" is used to determine the boundary elasticity in SVMs. The larger the “C” hyper parameter is selected, the greater the flexibility in the boundaries. In order to generalize the model, “C” should be chosen at a level that can make a small number of violations [24].

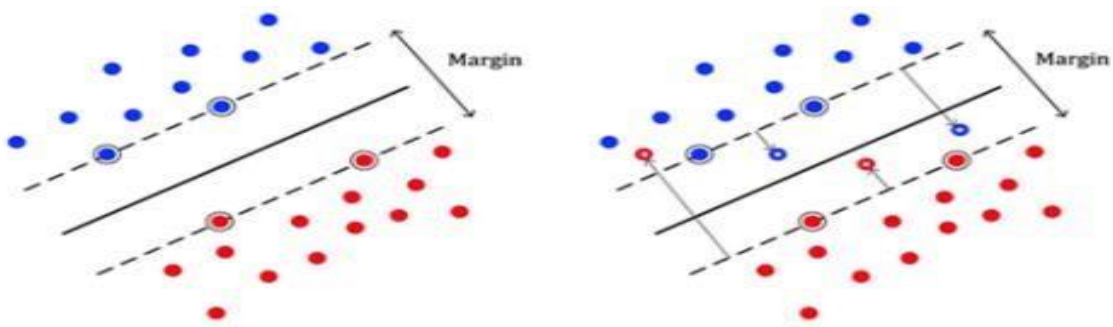

Figure 2.9: SVM Process 2.

In everyday life, no event develops linearly. Therefore, when the problems caused by such events are transferred to the machine environment, they cannot be displayed in a linear plane most of the time. In non-linear support vector machine classification, features can be added to display the data set clearly. This provides a linear distribution of data. There are some methods that can be used when classifying data that cannot be represented linearly with support vector machines. The most commonly used method is called Kernel Cheat. Although the method of adding features to non-linear data does not cause problems in small datasets, in complex and large datasets, it causes too many highorder features, that is, the number of dimensions to increase. This slows down the performance of the system and reduces the processing speed. The Kernel Cheat method was developed for this reason. The method called polynomial kernel is a sub-method of the kernel trick, and it makes it behave as if it was added to the model without adding polynomial attributes. Thus, the model that does not grow in size but acts as if it is, can make a successful classification, while not slowing down its performance. In the model using the polynomial kernel, if the model is overfitting, the polynomial degree is reduced, on the contrary, if the model is underfitting, the polynomial degree should be enlarged [25].

e. Linear Relation, the regression method, which is used in many fields from medicine to mathematics, from social sciences to statistics, is an analysis method that examines the relationship between two or more variables with a cause-effect relationship and models it with machine learning. It is divided into two as linear regression and logistic regression, depending on whether the variables used have a class or a numerical value of 39. Linear regression (linear regression) is the ability to express the relationship between dependent and independent variables with a linear line. If we want to formulate it mathematically, we get the following results. The chart above shows the selling prices of the houses, which vary according to the square meter. On the x axis there is the number of square meters, on the y axis there is the sales price, that is, the label values of the model. When a linear line is drawn according to the labels corresponding to each input value, we have a linear regression model. By looking at this model, a corresponding continuous label value can be found for an input value. To measure the performance of the model, the difference between the estimated tag value and the actual tag value should be taken. This is the residual value in the formula. In machine learning, it is expressed as a residual value loss function. The most commonly used method to find the missing value in linear regression problems is the Mean Squared Error

(MSE) method. 40 Mean Squared error (MSE) looks at how far the regression graph is from points included in the training. It does this by squaring the difference between the estimated value obtained from the regression line and the target value. The reason it is squared is to remove the negative sign. It is called the average because it performs this operation between more than one estimated value and the actual value. The closer the MSE is to zero, the less the error margin of the model and the higher the performance. The higher the MSE, the lower the performance of the model. The Mean Squared Error formula is as shown below [26].

f. Logistics Coupling Logistics regression is a type of supervised machine learning used in classification problems. Linear regression is not a quantitative output, but gives the probability that the given input belongs to a particular class. For example, a new message in an inbox has a ten percent probability of being a spam message. In the logistic context, as in the linear correlation, the result is produced based on the given input values. There are some differences between linear and logistic regression. While continuous variables are used in linear regression, discrete variables are used in logistic context. Logistic regression

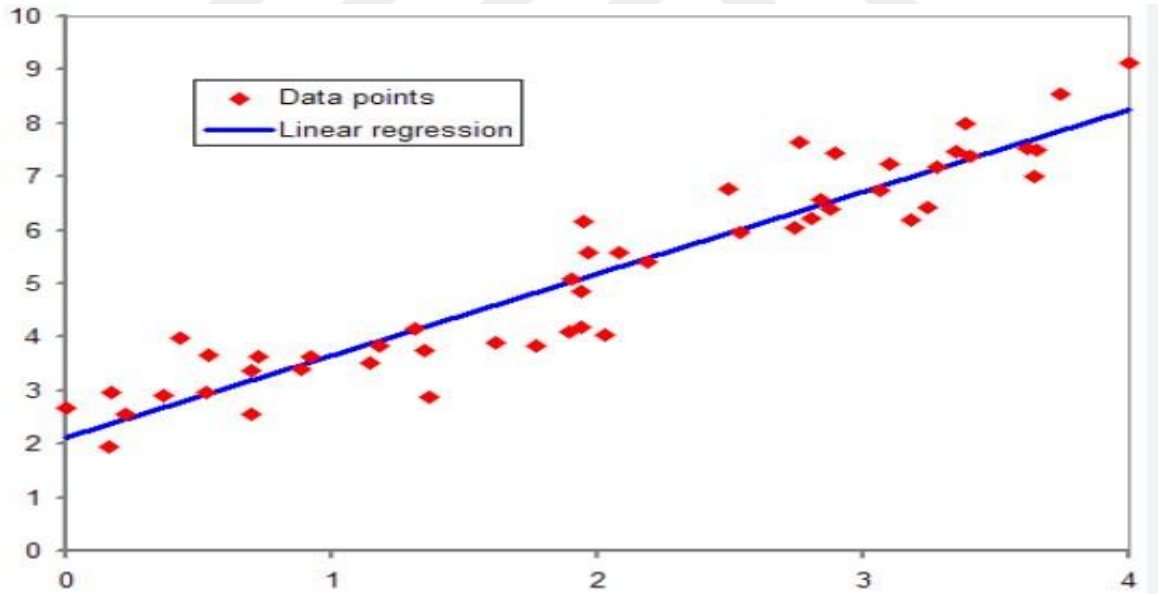

Figure 2.10: Regression.

assumes that the outcome variable is a random event. For example, if we analyze people by body fat ratio, the result may be obese or normal weight. Logistic regression considers the probability of people being obese. If the body fat ratio is above thirty percent, the person is

considered obese; otherwise, he is of normal weight. The mathematical formula of the logistic linkage is as shown below [27].

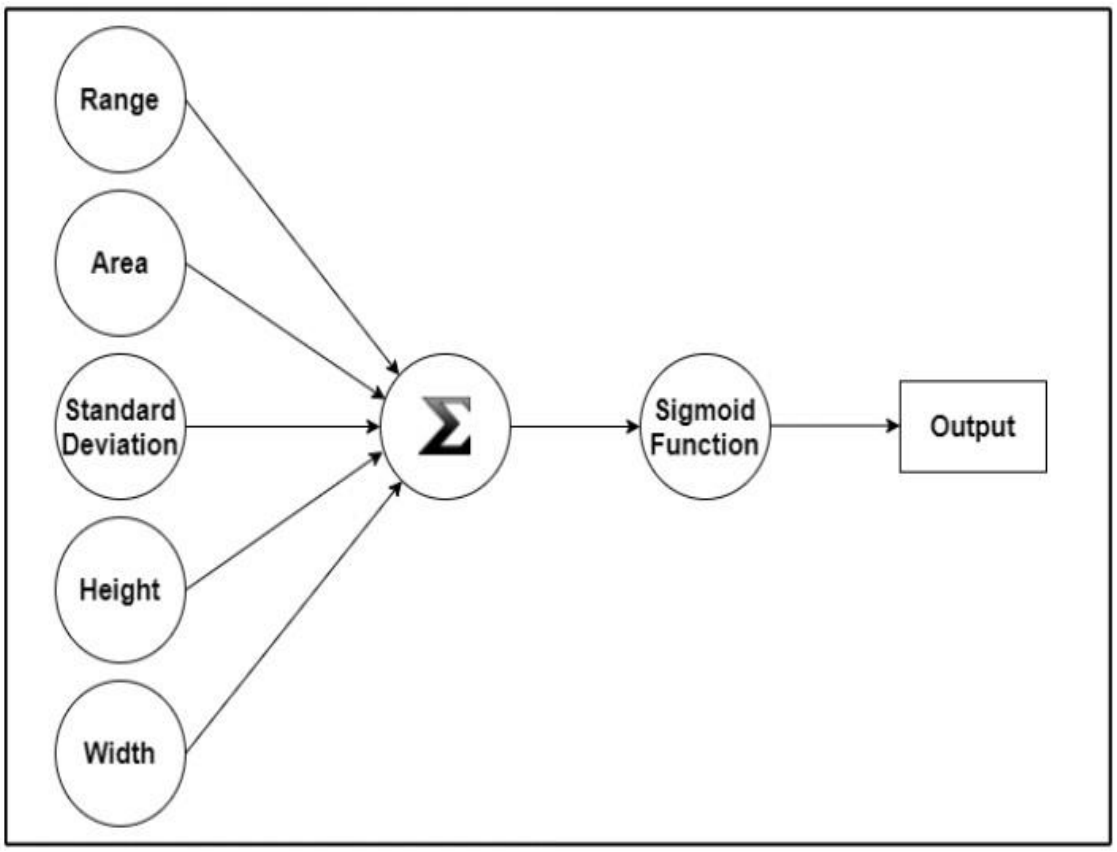

Figure 2.11: Logistics regression.

g. Artificial Neural Networks Artificial neural networks (ANN) method is a research subject that enables machines to gain human-specific abilities such as learning, making predictions, drawing conclusions and generalizing, modeled on the human brain nervous system as an example. Artificial neural networks, developed by referring to the working principle of neurons in the nervous system, aim to realize the human brain on a machine with the network structure they have created with the features of processing the data, transmitting it to the next neuron, and doing them in parallel. It can show the elements that ANN corresponds to in the biological nervous system as follows. The synapse establishes the connection between two neurons and enables the information received from the axon to reach the dendrite of the other neuron. Synapses are the part where the learning process takes place. The dendrite is located at the end of the neuron. It provides information from other neurons. The nucleus transmits the information it receives from the dendrites to the axon. The axon transmits the data

transmitted from the nucleus to a new neuron cell as output. Artificial neural networks consist of five basic elements. These are the input, weight, non-addition function, activation function, and the final output, respectively. Each input is multiplied by its corresponding weight, as shown in the following figure. It is then summed by adding the threshold value b. The sum of the function may vary depending on the function selected by the user (such as sum, product, maximum, minimum, etc. functions). It takes the data output value passed to the last activation function. Artificial neural networks are divided into two as single and multilayer models according to their layers. Single-layer Neural networks consist of only input and output layers. It is the simplest network structure. All inputs are combined in the output layer. In the single-layer network, which was referred to as the Perceptron model in 1960, the activation function gives a result of 1 if the value obtained from the sum function is greater than the threshold value determined by the user, and 0 if it is less. This is because it uses a step function as an activation function. In other words, the Perceptron model represents definite yes and no situations in binary classification. This model is based on learning on error. As the model encounters an error, it learns and reduces the error rate. The learning method used in the Perceptron model is the Hebb method. According to this algorithm, weight updates are made according to the following formulas according to the false 0 and 1 cases obtained in the output values. The model with another single-layer perceptron is the Adeline model. In the Adeline model, which is the abbreviation of the words Adaptive Linear Element, a different method called Delta is used, unlike the Hebb learning method used in the classical Perceptron model. In the Delta learning method, also known as the gradient reduction method, the difference of the output value obtained as a result of the linear activation function with the value planned to be obtained at the beginning, together with the product of a determined learning coefficient, means that the network weights are updated and thus the learning process is carried out. If we want to explain the delta learning method mathematically, the following results are obtained. Adeline and classical Perceptron models can be used according to the tasks requested from the model. While the Perceptron model using the step function as the activation function shows higher success in binary classification problems, the Adeline model using the sigmoidal activation function can show how far or close the result is linearly to the target value since it can produce a continuous value between 0 and 1. The following image shows the activation functions used in single-layer sensors. Since models with single-layer perceptron are

insufficient in solving nonlinear problems and only give two-class results, the subject of multi-layer models has been started to be examined. Multilayer neural networks contain hidden layers in addition to the input and output layers. The information received by the input layers is transferred to the hidden layer without any processing. Problems that cannot be solved by single-layer networks have been the subject of multi-layer neural networks, since real-life events are non-linear, that is, their results are not separated by a straight line when shown on a graph. The first problem in this context is the XOR problem. XOR is a logical operator. It produces 0 if the values given as input are the same, 1 if they are different. XOR is the first problem successfully realized by multilayer networks because the 0 and 1's cannot be separated by a linear line. Multilayer networks form the basis of current deep learning algorithms. Inputs related to the result desired to be analyzed and outputs that these inputs are intended to produce are presented to a MCA model, which consists of the input layer, the number of middleware and output layers specified by the user. For each training epoch result, the network takes the difference between the self-generated result and the targeted result and starts updating the initially randomly determined layer weights [28]. This is called error backpropogation. As with other systems, its purpose is to minimize the error. The error is calculated separately for each cell in the output layer. The total error is found by summing the errors calculated for each exit point. The part up to this point is called forward feed (forward propagation). Feed forward is part of backpropagation. Back propagation starts after the feedforward ends and the error is calculated for each node in the output layer. The model, which calculates the loss function obtained from each epoch result, that is, the difference between the obtained value and the expected value, and updates its weights, increases the success rate of the network by, enabling the system to improve its learning and gain the ability to generalize. Forward and backward propagation can be mathematically explained as. ANN, which differs from classical programming methods thanks to its ability to learn, predict and generalize based on old knowledge, forms the basis of new generation deep learning methods and is used in many areas from technology, finance, geophysics to health sciences and marketing with developing computer hardware. Artificial neural networks, the use of which is increasing with the increase in modeling for the solution of nonlinear problems, is important in reaching engineering-based issues such as examining

linear or non-linear systems, image processing, voice or speech recognition [29].

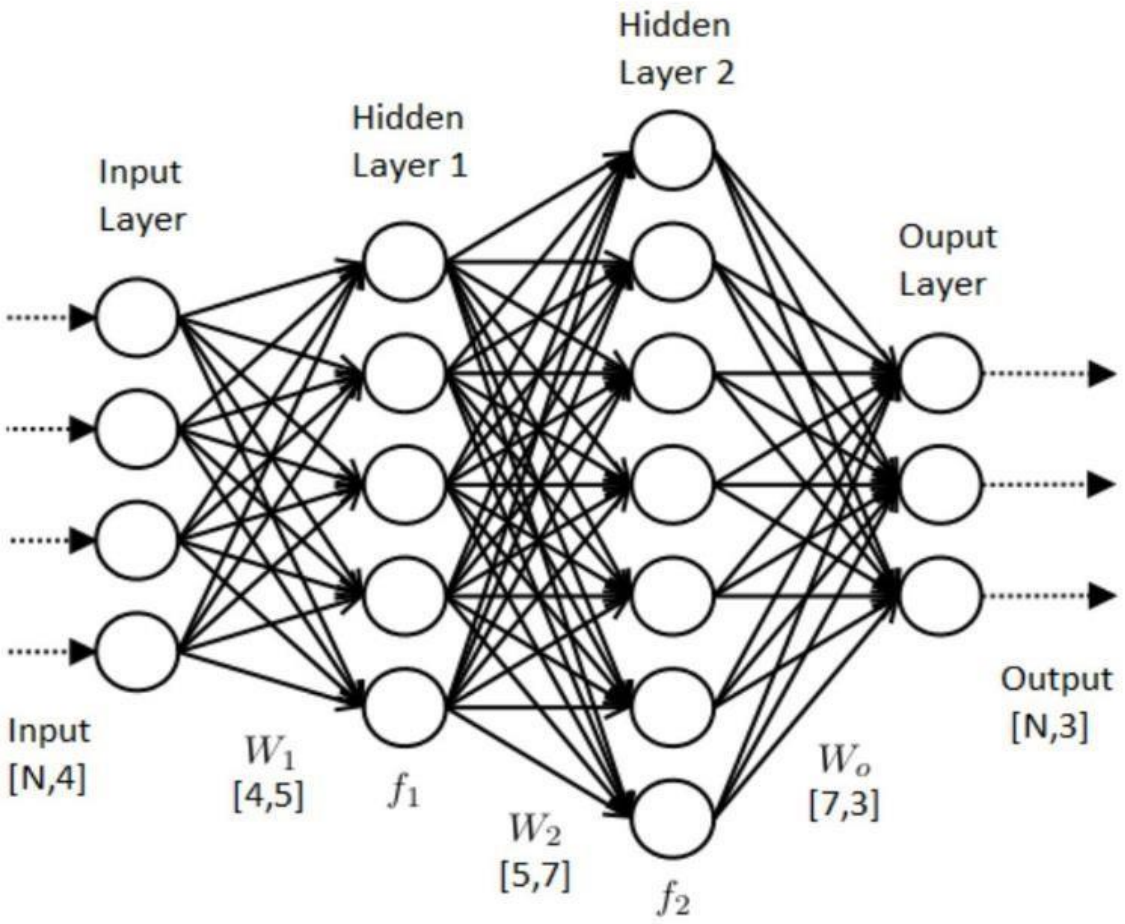

Figure 2.12 ANN.

2.6.2 Unsupervised Learning

In unsupervised learning, if the model is trained, the data is unlabelled. The model tries to learn by itself, to find similarities between data. To illustrate, let's say you have a dataset of visitors to your LinkedIn profile. The model can generate assumptions by grouping similar visitors. These assumptions may be that 50% of visitors are female software developers who have worked for 510 years and look at your profile in the morning, or that 30% of visitors are companies that need employees and look at your profile during working hours. This is called clustering in unsupervised learning. Separating the obtained clusters into smaller clusters is called hierarchical clustering algorithm [30].

Another unsupervised learning task is size reduction. The task of size reduction is to show the information in the simplest way without losing data. This can be done by combining several interrelated attributes. For example, the similarity between the age of a working civil servant and the seniority of the civil service may be high. The size reduction algorithm can treat these two features as one feature and act accordingly. This is called feature extraction in machine learning. The basic working principle of unsupervised learning algorithms is based on representing the data in a useful way. That's why visualization is so important. In models fed with data sets formed by a large number of heterogeneous data, visualization algorithms draw multidimensional representations to indicate each subset. For example, in the representation of two different sets that overlap in two dimensions, the third dimension provides the representation of the sets separately from each other. Another unsupervised learning algorithm is anomaly detection. When the model trained in normal situations sees another sample, it can detect anomaly according to whether it is similar to normal situations or not. For example, in the case of a credit card purchase more than usual, the system can detect this and accept it as unusual and warn the cardholder. The most used unsupervised learning algorithms in the literature, K means, PCA and others are shown below. The K-Means (K-Means) method is an unsupervised type of machine learning that is a clustering algorithm. It is a clustering algorithm that can output unlabeled data, that it does not know the output, by making simulations among themselves. K represents the number of clusters determined, 17. K center points are determined randomly in the data set and the distances of the data to these K center points are measured. The data belongs to the cluster with the shortest distance. Since the center points are chosen randomly, this process must be repeated several times to reach an optimal value. The purpose of finding the optimal clusters is that the elements of each cluster show the maximum similarity to each other and the clusters show the maximum distance from each other. This distance represents increasing the performance and reducing the similarity between classes when outputting because the labels of the data are not known. The elements of the K-number clusters determined in the K-means method should have a minimum variance for themselves [31].

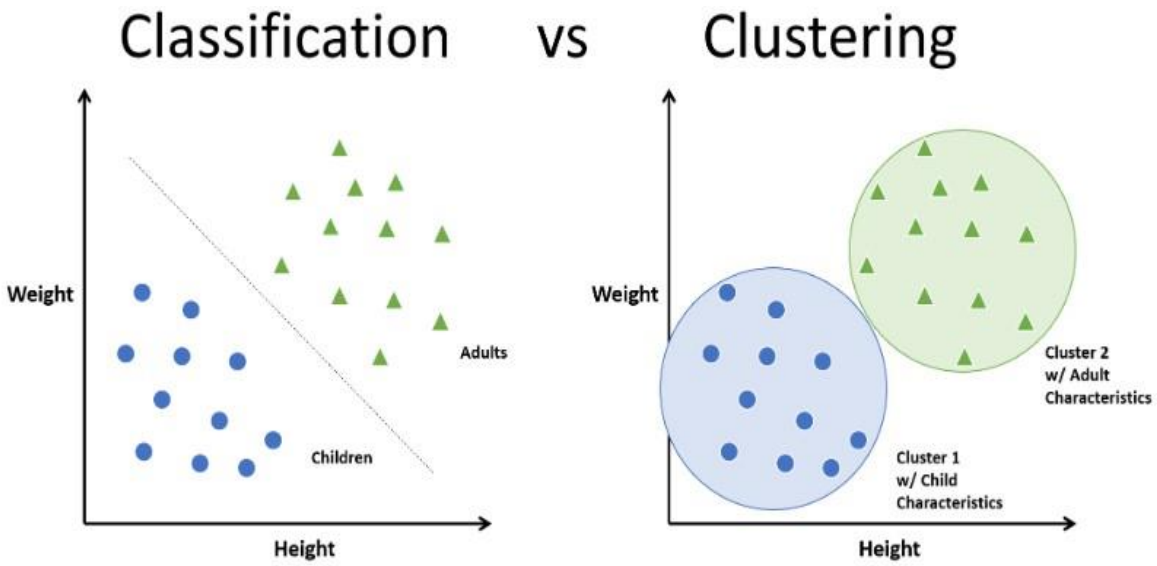

Figure 2.13: Classification and Regression.

In the K-means method, there are two different methods to find the appropriate K cluster value. These are Elbow Method and Silhouette Analysis. In the elbow method, the sum of the square of the distance from the center of the cluster is calculated for each number of K clusters of unlabeled data. A graph is drawn according to the calculated distances and the number of K clusters, and the value in which a sudden decrease is seen in the graph gives the optimal K value. In the silhouette analysis method, the following equation is applied according to the number of K clusters. According to this equation, the silhouette coefficient formed for each element in the clusters takes a value between -1 and 1. A value of 1 indicates that the data belongs to the cluster from which it was calculated. A value of -1 indicates that the data is grouped with an incorrect cluster. As seen in the graphic above, for the most suitable clustering method for this data set according to the Silhouette method, the K value of 5 with the maximum silhouette score should be selected. Thus, the optimal K value is found for this data set. Principal Component Analysis (PCA) method is an unsupervised learning algorithm that is generally used in multidimensional datasets for reasons such as noise reduction and size reduction. The purpose of principal component analysis is to purify the multidimensional data set from features that can be called more detailed. In other words, the aim is to ensure that there is no information leakage from the main data when the size of the data is narrowed. It is essential to be able to apply this to less important attributes in the data, since it will be impossible not to lose the data whose size is reduced. When it is thought through a graph, being able to look at a complex data with multiple features from a different

perspective and to make more meaningful distinctions between the data is called PCA. As can be seen in the image below, looking at a large number of complex datasets from different points causes a reduction in size [32].

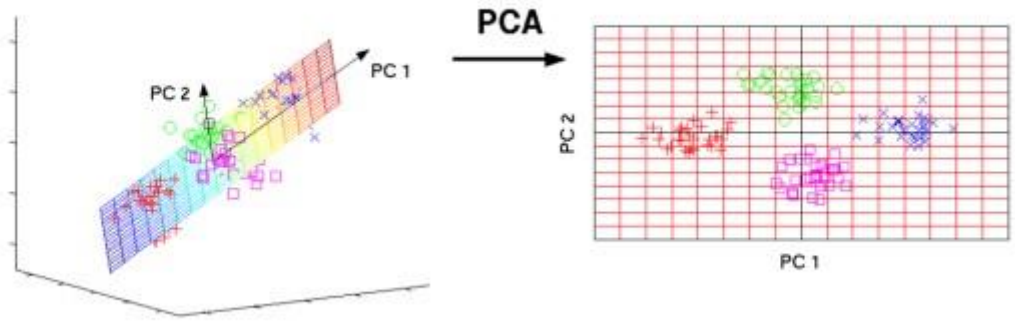

Figure 2.14: PCA.

Principal component analysis is to ensure that the raw data in the i dimension is expressed in k dimensions, without losing its essence, which is its most meaningful feature, in i<k. The variables in the k-dimensioned data are called principal components. The first transformed principal component is the component with the largest variance, and the other transforming components continue with decreasing variance. The variance in a random distribution is the square of the difference between the variable and the mean of the distribution. The reason for wanting to have a high variance is to display unlabeled data in a way that can distinguish them from each other. Eclat is an unsupervised machine learning algorithm based on association rule extraction. Association rule inference is a method that finds the probability of occurrence together by solving the similarities between the data. The most popular example to be given to this method is the systems that make suggestions according to the products that users have bought or viewed before, in online shopping sites that allow shopping over the internet. In other words, the association rule inference method examines and learns the behavior of the users, making suggestions about the next behavior of the users or the sellers as shown in the example above, about which products are sold together. There are two important points in the association rule extraction method. These are the Support and Trust factors initially set by users.

The Eclat algorithm is a working method given in vertical format, not in horizontal format, unlike Apriori, which is another association rule extraction algorithm, which informs its

users about which products are used together or which data is close to each other. It finds frequently used items in the master dataset in Eclat. This is called deep priority search. When we want to solve the example given in the table above with Eclat, the following steps must be done: First of all, the data given in the horizontal format should be converted to vertical format. TID Items 1 Bread Butter, Jam 2 Butter, Soda 3 Butter, Milk 4 Bread, Butter, Soda 5 Bread, Milk 6 Butter, Milk 7 Bread, Milk 8 Bread, Butter, Milk, Jam 9 Bread, Butter, Milk 22 elements should be discarded according to the minimum support value (minsup) selected by the user. In the example whose steps are shown below, this value is 2. There will be no discarded item, as there is no oddnumbered transaction number list. After the field, as a second step, the frequencies of the items in the basket will be found in the form of binary combinations. From this list, the row with the only remaining transaction frequency will be destroyed. Then triple combinations of items will be made and this process will continue until all the number values in the TID column are less than the minimum support value. For example, when the minimum support value is selected as 2, when all the values in the transaction number list are 1, it stops and the suggested combinations in the table obtained in the previous step are started to be used [33].

2.6.3 Semi-Supervised Learning

Semi-Supervised Learning It is a form of learning where labelled and unlabelled data are combined. In this type of learning, both supervised and unsupervised learning algorithms work in a hybrid way. They usually work with mixed data with lots of unlabelled and few labelled data. They can match the goals they learn from labelled data, that is, possible outcomes, with clusters in unlabelled data. Apps like Google Photos and Pinterest are examples of semi-supervised learning.

2.6.4 Reinforcement Learning

Reinforcement Learning is a technique of machine learning and differs from supervised and unsupervised learning in terms of its working principle. Reinforcement learning is an area that is very different from other machine learning systems and is continuously studied separately. As in other types of learning, the subject is based on the sequential behaviour of the study process according to the reward and punishment system, not the classification of the input data according to whether they are labelled or unlabelled, or the clustering of

similar data in the most appropriate way. In Reinforcement Learning, the learning of the model is reinforced by the system unit called the agent, in a continuous learning action, through the application of reward, that is, positive, for actions that it considers to be correct, and punishment, that is, negative feedback, for actions that are wrong. That's why it's called Reinforcement Learning. In this technique, the set of actions determined by the agent, namely the system, that is, the algorithm chosen is called policy. Policy is the agent's responses to the situations he encounters in the environment where he will perform his task. The unit called the agent is not considered as an objective entity, but as a tool that enables to perform actions and learn. Initially, randomly determined actions will adapt to the environment in which the training takes place and begin to make logical decisions as the model is trained with rewards and punishments. The purpose of Reinforcement Learning is to ensure that the system reaches the maximum level of reward after the learning process or the minimum penalty point in cases where it can receive continuous negative feedback. The reward expression is associated with the concepts of pleasure and pain in reinforcement learning systems and serves as a feedback signal. Based on these feedbacks, the agent can make a choice based on which of the actions will receive more positive responses, by making use of their experiences. In some cases, the opposite action can be chosen instead of choosing the action with a high score or a low penalty score instantly from the problems faced by the agent. This is because other actions linked to the momentarily disadvantageous action have higher positive or lower reward penalties. This is called the state value in reinforcement learning. It is considered as possible total reward or penalty points, not instantaneous. In reinforcement learning systems, whose purpose is to optimize rewards, policy selection has a very important place in increasing performance. Choosing a dynamic policy to meet all possible outcomes during the mission is crucial in optimizing the reward the agent can obtain. Reinforcement learning systems can be used independently of the environment, called a model, without a model, and, unlike the model, they can perform without being trained beforehand [35]. It is frequently used in desktop and computer games, smart home appliances in terms of learning by trial and error method, where reinforcement learning experience is important. Contrary to supervised and unsupervised learning, reinforcement learning, which takes place in situations where the training set is not used, is carried out by the agents performing and learning the task more than once. This is because the agent can recognize the task due to the large number of uncertain conditions and sub-conditions.

Reinforcement learning is likened to the decision-making and learning system of living things, especially animals, in terms of the absence of a training data set, that is, making decisions without the need for any preliminary data. There are some reinforcement learning methods that have been developed to increase the decision-making process and the accuracy of the decisions taken. The most widely used of these are Markov decision processes and QLearning method. Markov decision processes, defined by Richard BELLMAN, are used to mathematically formulate the decisions that the agents who perform the axioms in reinforcement learning systems must make in order to move from their current state to the next state. According to the Markov decision processes, the transition of an agent in the St state at time t to the St+1 state depends only on the St state. This rule is called memorylessness of Markov decision processes because the decision process depends only on the current (t) state, not on the past states. For the agent who receives reward or punishment according to the axioms he makes and tries to optimize it, his transition to the next state depends on the probabilities of one or more of the axioms he needs to decide.

3. MATERIAL AND METHODS

3.1 ARTIFICIAL INTELLIGENCE (AI)

In this section, maturing technologies of 4.0 are explained in order to reveal the benefits of technologies that support change and innovation in many sectors, drive the fourth industrial revolution and enable the digital real estate revolution, and the applications that can be associated with them in the real estate sector.

Digitization can be defined as the name given to the process of transferring accessible data to digital media so that it can be used by technological devices (computer, tablet, phone, etc.), arranged in digital environments and included in the working process. Digitization is the process of transforming existing data into new products to be strengthened by using communication technology infrastructure for companies to create new working methods, obtain innovative outputs and effectively benefit from all assets held by companies, that is, making data more usable by digitizing data with technological infrastructure. Digitization is the use of digital technologies and applications with the aim of creating new value and opportunities to develop or revise the working model. The impact of digitization activities on the real estate sector, as in all sectors, will continue to increase day by day. The main technologies of digitalization, whose impact will increase day by day in all areas of real estate, including the construction, use, operation and management of buildings and capital markets, are as follows [36].

3.1.1 Cyber Security

Businesses that use cyber/digital systems and high-speed connections in order to provide efficient customer service and maintain economic management services are using cyber security systems to secure their digital assets and protect their systems from unwanted access. Cyber security refers to all attempts to establish a secure data processing area in order to ensure that the integrity of the data is not compromised during the storage and transfer of data or information in the digital space and that unauthorized persons cannot access this data. Cyber security, also known as information technology security or electronic information security; are applications designed to protect electronic devices, networks and data against

malicious attacks. Security has emerged as an important concept since the emergence of the discipline of international relations.

In this framework, both states and many international organizations have developed policies based on security. The concept of security has been defined in different ways in different periods. It is a concept that is defined differently from person to person, from group to group, from state to state. Security is a concept that has expanded from a single person in life to forms of social organization such as society and the state. In Abraham Maslow's human need hierarchies, the need for security was seen as the second important step after food and shelter. Security has been defined as a state of feeling free from fear, danger and threats. From this definition, it can be easily understood that security has both physical and psychological dimensions and is open to subjective and objective definitions. In particular, the concept emphasizes the subjective side of security with its mental processes in terms of 10 psychological and emotional dimensions. As a matter of fact, the subjective security situation is described as bringing the fear of attack on values to the fore and the absence of these attacks. If the emphasis is placed on subjective security, the emphasis on the threat situation and the measures and precautions against the threat gains more importance. In an objective security definition, only the absence of threat against the values gained is stated, while the emphasis on the measures and precautions is not as intense as the other. Throughout history, the physical dimension of security has been emphasized, and it has been seen that the borders of political organizations such as the state correspond to the understanding of keeping them away from the attacks and threats of other states. A state or the community living on it feel safe; It has been a subjective phenomenon with a physical and mental dimension, corresponding to the elimination of the possibility of insecurity. This situation is also defined as seeing insecurity situations as a threat. The concept of threat, on the other hand, could be shaped according to the possibility of realization on the basis of perceptions and predictions, as well as being based on real facts and events. It is clear that in order to talk about the security phenomenon, internal or external threat perception and predictions are needed for the protection and maintenance of the security asset. Logical processes available at the individual level for security; Reproduction in organizational forms such as society and the state has brought about the issue of security at different levels. The emphasis on collective security in the field of social life until today has been mostly within the framework of the national security of the states, and the state has been considered the main actor. Cyber

security can be defined as a set of tools, policies, security concepts, security guarantees, guidelines, risk management approaches, activities, training and technologies used to protect the assets of institutions, organizations and users in the cyber environment. The concept of cyber security should be divided into two and defined. The concept of cyber the word cyber is a word used to describe concepts or entities that are linguistically related to or involve computers or computer networks. Again, the word cyber space, which is mostly used in the literature, is used to describe the abstract or concrete space in which interconnected hardware, software, systems and people communicate or interact. The word, which is expressed as cyber in terms of linguistics, is also used in the meanings of "belonging to computer networks", "belonging to the Internet", and “virtual reality" by referring to the word "cyber" in English, and today it is a concept that expresses the space formed by information and communication networks. Appears as a concept. The word cybernetics, on the other hand, is derived from the word cyber, and was first described as "control and communication in animals and machines" in Norbert Wiener's work "Cybernetics". The increase in computer use has led to unprecedented social changes and has begun to offer important conveniences for human life in daily life. With the widespread use of the Internet, geographical borders have disappeared, and a new world, defined as the "cyber world", has emerged that enables people to access all kinds of electronic information services. Cyber can be defined as a global system, generally known as a network that hosts computer systems, communication infrastructures, databases and information tools connected to the Internet. Based on these definitions, we can conclude that cyberspace today does not consist of a single space, but is an intertwined space, each of which provides different digital interaction and communication methods. We can talk about the existence of a rapidly developing and inhomogeneous global system [37.

3.1.2 Cyber-Physical Systems

Cyber-physical systems, which are defined as two elements that combine real and digital elements as a whole, can operate at different spatial and temporal scales, exhibit many and various motion patterns at the same time, and communicate in different styles according to conditions, are a physical system. They are systems in which the object is controlled or monitored by software created on technological devices (US National Science Foundation, 2010). Cyber-physical systems is the connection of the real world with the digital world

through components that usually include coding systems, communication technologies, sensors, along with embedded technologies. To give an example of SFS; projects such as smart network, autonomous vehicles and applications, industrial control systems, robotic automation systems and autopilot electronic systems can be shown [38].

3.1.3 Cloud Computing

Cloud computing, which is perhaps the most common of the technologies supporting Digital Real Estate; It is an application that uses a network of servers maintained on the Internet to store, manage and process data. This means that files previously held at work can now be accessed from any number of compatible devices by anyone with permission, anywhere in the world. Thanks to this technology, platforms are designed that enable real estate owners to manage their real estates in the simplest way possible. Further efficiency gains are also expected as Microsoft and RIB software companies develop a cloud solution for BIM [39].

3.1.4 Virtual and Augmented Reality

Reality technology is divided into virtual reality and augmented reality. While virtual reality defines an area that is used more in 3d games and where people's relationship with reality is completely eliminated, augmented reality defines an area that allows users to see the real and virtual objects together on technological devices, where data and images in the virtual environment can be added to the image at any moment in our lives. Combined with YBM, SG and AG, it enables architects to create an almost lifelike interpretation of their designs. With virtual reality technology, which is more advanced in terms of development compared to augmented reality, any building can be created in 3D in the virtual world before starting the construction, and by analyzing the labor-machine-equipment-structure relationship in this environment, errors arising from the complexity of the nature of the construction industry can be reduced or eliminated, and it is possible to go to the construction site for possible problems. can be resolved before it starts [40].

3.1.5 Sensors

The synonyms of this word, which was translated into Turkish with the adaptation of the English “to sense” dictionary, are “perceptron” or “sensor” (Yüksel, 2006). Sensors detect and calculate any stimulus (physical or chemical parameters such as pressure, electricity,

heat, light, acceleration, humidity, sound, force, distance and pH) occurring in the external environment and turn them into electrical stimuli (Pohanka, Pavlis, & Skladal, 2007). Sensor technology is the toolkit that enables Digital Real Estate companies to save their data and increase their efficiency. With the development of smaller, cheaper and smarter sensors potentially embedded in other devices, further gains will be made for the real estate Blockchain Digital Currency Record Retention Security Smart Contracts 32 industry. This connection between devices and sensors of all kinds is currently referred to as the Internet of Things. Modern IoT Sensors, Temperature Sensors, Pressure Sensors, Humidity Sensors, Flow Sensors, Accelerometers, Magnetometers, Gyroscopes, Inertial Sensors, Image Sensors, Touch Sensors, Proximity Sensors, Acoustic Sensors, Motion Sensors, Occupancy Sensors, Image Processing Occupancy Sensors (IPOS) , Intelligent Occupancy Sensors (IOS), CO2 Sensors, Light Sensors and Radar Sensors can provide information on a wide variety of environmental indicators

3.1.6 Blockchain

Blockchain, put forward in 2008, is a directed record-keeping domain, in other words, a directed, shared, coded, non-refundable and unchanging data collection domain. Blockchain is a database that confirms and stores all the mutual activities of those who benefit from the model through the network, and the formation of the blockchain can be summarized as follows; Transactions kept in blocks are connected to each other and written to the system by forming a chain, then the block is spread to all distributed registers using the current communication network and added by confirming with coding. The summary of the previous block is taken and a new block is created so that the next block is produced and added to the chain, and then each node confirms this transaction by any two users in the system and keeps its record, which confirms the block, which ensures that no one else can change them, and the structure of the system is maintained in this way. Although blockchain is a technology that has started to be known with the finance sector, namely Bitcoin, this technology, which has been used in many sectors in a short time, has provided instant access to the title deed and land ownership data of the relevant stakeholders and agencies, enabling the asset owner to reach the documents they need, and providing easier information about the property to the concerned. It has been ensured that both cost and time losses to individuals and institutions have been prevented [41].

3.1.7 Websites and Smartphone Applications

User interface between suppliers and customers; all websites, including social media sites and mobile application sites, are prepared to increase the efficiency of the digitalization process by focusing on user experience, collecting data and analyzing data. According to mobility data provider GYANA, a smartphone sends location data to a nearby cell tower about 20 times per second, and many of us present aggregated data in this way, even though the law prevents citizens' personal data from being used against their will. We transfer most of our data protection privileges after accepting the terms and conditions when downloading applications and using online services. Mobile location analytics provides understanding of consumer behaviour and the use of smartphone data helps to understand the behaviour of individuals in the city and creates new smart urban systems [42].

3.1.8 Application Program Interfaces

To increase the efficiency of the real estate market, Digital Real Estate applications need access to the data needed. An API or API is a set of functions and procedures that access the properties or data of an operating system or application. Open access APIs enable real-time aggregation of real-time data from different sources without huge implementation costs, but most of the realestate data needed is privately held and contained in analog documents, so it is clear that this data is not that easy to collect or access.

3.1.9 Big Data Analysis

Big data is defined as the use of new technological storage, processing and usage methods when needed, when the size of the data used for operation/application purposes reaches very serious levels. Big data is defined as the original analysis, processing and storage of big data sources, which are mostly gathered from various data sets and where traditional data analysis methods are not sufficient. In other words, big data enables many needs to be met by bringing together many independent data sources, compiling information that does not have much connection with each other, and bringing together unknown information in a very short time. Those who use digitized systems do so mostly through generic software; 60% of executives use spreadsheets as the primary tool for their firm's reporting, 51% for valuation and cash flow analysis, and 45% for budgeting and forecasting. This situation prevents access to 'big

data'. By increasing access to information, data science supports project managers in making more effective decisions. Big data for real estate; social media activity, trip advisor reviews for any neighborhood, phone location data, etc., produced in near real-time and used to interpret traditional regression and spreadsheet models [43].

3.1.10Internet of Things (IoT)

Internet of Things (IoT); It is a network structure in which technological devices, devices connect with each other and collect information by sharing data without the need for any human influence or input, and decide with the data obtained. Another definition is the Internet of Things, where addressable objects develop with each other and these objects communicate in accordance with various protocols in a global wide network. In addition, the Internet of Things is defined as any device that can connect to the Internet, and it is estimated that the number of these devices will reach 29 billion by the end of 2022. Data collected by a large number of individual electronic devices (sensors, switches, light bulbs, phones, cameras, refrigerators, etc.) informs smart buildings and ultimately smart cities, creating big data so that advanced analytics can be prepared. IoT also enables the development of Building Information Modeling (BIM) technology, which is a digital simulation or model of a real estate. Because BIM, driven by IoT, which is increasingly used, will ensure that a building is used more and more throughout its life cycle [44].

3.2 DEEP LEARNING

Deep learning has emerged due to the slow and ineffectiveness of machine learning algorithms in solving the problems caused by the amount of data growing with the developing technology and the technological equipment advancing in direct proportion. The inadequacy of machine learning methods with billions of data, mostly images, used in social networks and other environments used over the Internet, is one of the reasons why deep learning has been the most used artificial intelligence sub-branch in recent years. Although machine learning algorithms are mostly based on mathematics and statistics, the fact that deep learning is more suitable for programming provides a great advantage in solving problems. Deep learning is based on the logic of artificial neural networks explained in the previous sections. The Multi-Layer Perceptron (MLP), which is processed sequentially in

many hidden layers and which results in the output layer by transmitting a hidden layer output as an input to the next hidden layer, is the working principle of deep learning.

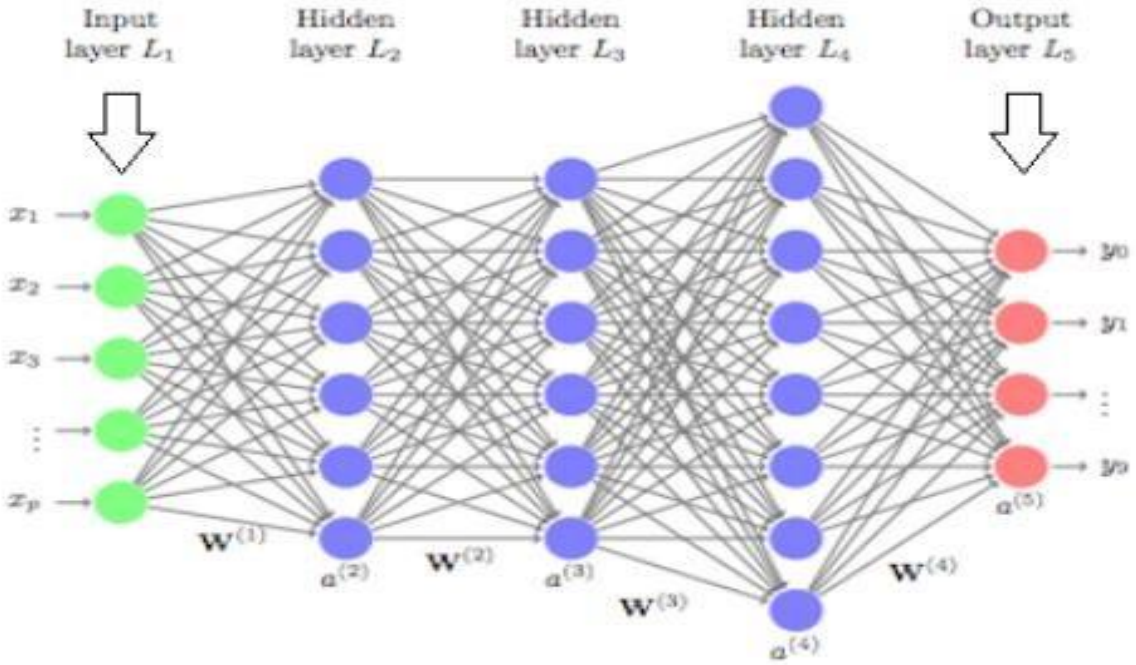

Figure 3.1: Deep Learning.

In the deep learning method, the data given as input is learned from the beginning to the end, from the more basic features to the features that can be considered more detailed. In deep learning, which is called deep due to this hierarchical learning between layers, the number of depth depends on the number of layers created. In order to achieve a high performance in a deep learning model, there are some operations and important points to be considered, just like in a machine learning model. These can be listed as preventing over- or under-fitting of the model, ensuring the optimal tuning of the hyper-parameters, choosing the appropriate optimizer, diversifying the data set if the data is needed, and choosing the validation method appropriately. When considering hyperparameters for a typical machine learning algorithm, they can be listed as batch size (batch_size), type of optimizer used, and learning rate [45]. For Deep learning with a multi-layer structure, the number of hidden layers to be used in addition to these and the activation functions of these layers are also important in hyper parameter setting. Hyperparameters are the units whose purpose is to reduce the loss function, to show a better performance of the model, to generalize the model to the data it has not seen before and to give good results. Gradient Descent method can be shown as an

example to show the importance of adjusting the hyper parameters. Gradient descent method, which is the most widely used machine learning technique among the optimizers whose aim is to reduce the loss function to the smallest possible value, is updated with each iteration and the loss function (cost function) is tried to be reduced to the global minimum, that is, to the lowest possible value. In this method, in each iteration, in order to find the weight value in the next iteration, the derivative (slope) value at that point is multiplied by the hyper parameter called the instruction rate and subtracted from the current weight value. The gradient descent formula is as shown in the equation below.

The gradient descent method can be thought of as the rock left from the top of a mountain system reaching the lowest part of the mountain. Here the lowest part represents the global minimum point. The learning rate can be thought of as the oscillations made by the rock. It is undesirable for the rock dropped from a height to remain in the so-called local minimum while trying to find the global minimum. The working principle of the gradient descent method is as shown in the image below.

3.3 LEARNING RATE

Learning rate is one of the units used in hyper parameter setting. It determines the step interval of the model trying to reach the global minimum. In models with low learning rate, the time to reach the minimum value of the loss function will be long and the model will have to make more iterations. On the other hand, layer weights change more consistently and the resulting model will be more successful. In networks with a high learning rate, the difference between varying weight values is very large. Therefore, the loss function cannot reach the global minimum point, but the training time of the model is very short. High learning rate is a disadvantage in determining the loss function. For the reasons mentioned above, while trying to minimize the loss function while determining the learning rate, on the other hand, the training time of the model should not be ignored. Therefore, choosing an optimal learning rate between low and high learning rates is important for model performance. The effect of learning rate on finding the loss function is shown in the images below [46].

3.4 OPTIMIZER SELECTION

Optimizer selection is the part that performs the learning process in deep learning. It aims to update the weights of the layers according to the loss function after each epoch of the network and to increase the performance of the model. Together with the training rate and layer weights, they form the hyper parameter setting. Although it is appropriate for small datasets to try all the options to find out which optimizer is more suitable while creating the model, this method does not disadvantage the users in networks with hundreds of thousands of data and where the training period takes days. The most commonly used optimizers are Gradient based optimizers (Stochastic Gradient Descent, Gradient Descent, Stochastic Gradient Descent with Momentum), AdaGrad, RMSProp, AdaDelta and Adam optimizers used in this study. Gradient descent method, the basic working principle of which has been shown in the previous sections, has its own types. The first of these, Stochastic Gradient Descent (SGD), updates the network parameters, that is, the layer weights, over only one randomly selected sample at each iteration step. The concept of stochastic, meaning random, in its name is due to the random selection of the samples it uses. Although the normal gradient descent method produces more noise than the optimization method that works with all data at once, it is more advantageous to use compared to other gradient methods in cases where there is a large amount of data. Because in each iteration, it does not work with all data, but only on selected random data. This greatly shortens the training time of the model. Momentum and SGD is a method that reduces the noise generated by stochastic gradient descent while trying to reach the minimum loss value. The minimum value of the gradient descent algorithm with momentum added ensures faster convergence and greater gradient oscillations. Gradient descent formula with added momentum is as given in the equation below. The image showing the oscillation differences between stochastic gradient descent and momentum added SGD is as shown above. Mini Batch Gradient Descent, which is another gradient descent method, tries to extract an optimal optimizer by taking the advantages of the two methods, which are located between the classical gradient method and the Stochastic Gradient method. It neither takes a sample per unit time nor scans the entire data set at the same time. Instead, it divides the training set into mini-heaps and works on these heaps at each iteration. It contains a parameter called “batch_size” and this parameter is 32,64,128…. The mini-stack gradient descent method is used in deep learning as well as other machine learning methods [47,48].

3.5 OVERFITTING AND UNDERFITTING

Overfitting and underfitting problems are the most undesirable problems in machine learning models. They directly affect model performance. It is due to the fact that the model trained with the training data set cannot perform the learning process successfully when it cannot see enough data, or that the model memorizes during the training. The first is called insufficient learning and the second is called overfitting [49]. These two problems cause a good problem not to be retrieved in the test data. The overfitting problem is seen as the model's evaluation metrics moving in the opposite direction after the epoch value, where the performance metric reaches the maximum value, and the loss metric reaches the minimum value during training. The overfitting problem is the most common generalization problem in machine learning and deep learning, with developing technology and growing data volumes. Generalization is the ability of the model to perform its task successfully on the data it sees for the first time. In deep learning problems, the first measure to overcome the overfitting problem is to shrink the mesh. By network is meant the collection of convolutions, accumulation and fully connected layers, which will be seen in the next section [50]. With the shrinking network, it is aimed to reduce the number of parameters that will occur at each layer output. Another method is to add transmission damping, that is, noise, between layers. In this way, some of the learned information will disappear at each layer output and the model will not be able to memorize. Another method is regularization of layer weights. The purpose of this method is to prevent large changes in the layer weights arranged after each epoch. Insufficient learning is due to the fact that the model is trained with too much data to gain the ability to generalize during the training. In such cases where the data is not sufficient, a new and larger data set is obtained by tampering with the existing data (changing the color of the size, shifting the image, etc.) with data augmentation. Thus, the model is trained with enough data to generalize [51].

3.6 PROPOSED COMPUTER NETWORK TRAFFIC CLASSIFICATION METHOD

Convolutional neural networks are a special type of feedforward neural networks developed from artificial neural networks. The fact that they have updatable neuron weights and threshold values causes convolutional neural networks to derive from artificial neural

networks. The difference between the two network systems is that the input data type that the ESA architecture allows to use is visual data. Thanks to convolutional neural networks (ESA), the importance and use of processing this data has increased with the amount of data growing in parallel with the advancing technology. Since the encountered data can always be in the form of images, not in vector form, it cannot be processed with multi-layered sensors. When high-performance GPUs reach a level that can match large data, convolutional neural networks are activated and started to be used in cases where fully connected layers or multi-layer sensors are insufficient. Convolutional neural networks, like other multilayer sensors, assign the output value from the previous layer as input. The reason why fully connected layers are insufficient is that while converting the data into vector form and learning from there, it cannot learn from the attached data next to that data and cannot establish the relationship between them. On the contrary, in convolutional neural networks, learning takes place in a hierarchical manner step by step. With each convolution layer, pre-learned features are superimposed and more specific information is extracted, so that there is a relationship between data (between patterns). In this way, the network performs better in complex data, as it performs learning by overlapping in a systematic and regular way. This is called the learning hierarchy in convolutional neural networks. At the same time, in convolutional neural networks, the layer does not need to relearn the information (pattern) it has learned from an image data when it sees it in another image or when it encounters it again at different points in the same image. This leads to convenience and time saving in the learning process. In fully connected layers, when the network encounters a pattern it has seen before, it has to re-learn that pattern. This repetitive learning process imposes a burden on the network and therefore on the computer processor. The feature that distinguishes convolutional neural networks, and therefore deep learning, from other machine learning methods is that feature extraction occurs automatically in layers by itself. In convolutional neural networks, the layer where this feature extraction takes place is the convolution layer. The data obtained from this is called tensor. Then, the patterns emerging from the convolution layers should pass to the fully connected layer to realize the problem that is the purpose of the model, such as classification. In order to achieve this, there is a flatten layer that provides the transition between these two layers. A convolution layer also consists of successive convolution and pooling layers. In other words, convolutional neural networks

are examples of multilayer perceptrons. The image below shows a simple convolutional neural network model.

Convolution Operation The convolution operation, which gives its name to convolutional layers, is actually a mathematical operation. It determines how closely the two given functions overlap each other. The convolution process, which is also used for the same purpose in deep learning, creates output feature maps by using filters from the feature maps given as input. Filters, also known as Kernel, act as feature descriptors in convolutional neural networks. Each filter searches for a certain feature in the attribute map in the form of a matrix. These features can be specific features such as whether there is a red color, an animal, or a facial information in the given image. The filters themselves are also available in matrix form. They extract 59 parts by shifting them along the attribute map and produce a scalar value one by one. When the entire feature map is scanned, the scalar values produced are combined to form a matrix again. This matrix is the output attribute matrix of that input attribute matrix. While creating the output matrix from the attribute matrices, the elements in the matrix part compared with the filter are multiplied by their corresponding values and finally all the values are added. Thus, a single value is obtained for each shift step. The output map looks for the presence of the feature in the filter in different regions.

The pooling layer is located after the convolution layers. The reason for this is the size reduction with the help of selective permeability from the data coming out of the convolution layer. The pooling layer works in two different ways. The first of these is the average hoarding, the other is the best hoarding. Average accumulation advances the window in the feature map resulting from the convolution, takes the average of the elements of the window at the time it is found, and gives it to the layer output. The method of collecting the best ones is that the window that can navigate on the feature map outputs the largest of the elements it sees at that moment. Thanks to the pooling layer, there is a reduction in the size of the data, but since this model will work with smaller sizes, it saves the training time for the model.

The genetic algorithm applied to enhance and find the best parameters of pooling layers in the featues extraction stages. Furthermore, the size and number of parameters also estimated using the genetic algorithm. Then, the features classified to the normal and abnormal labels.

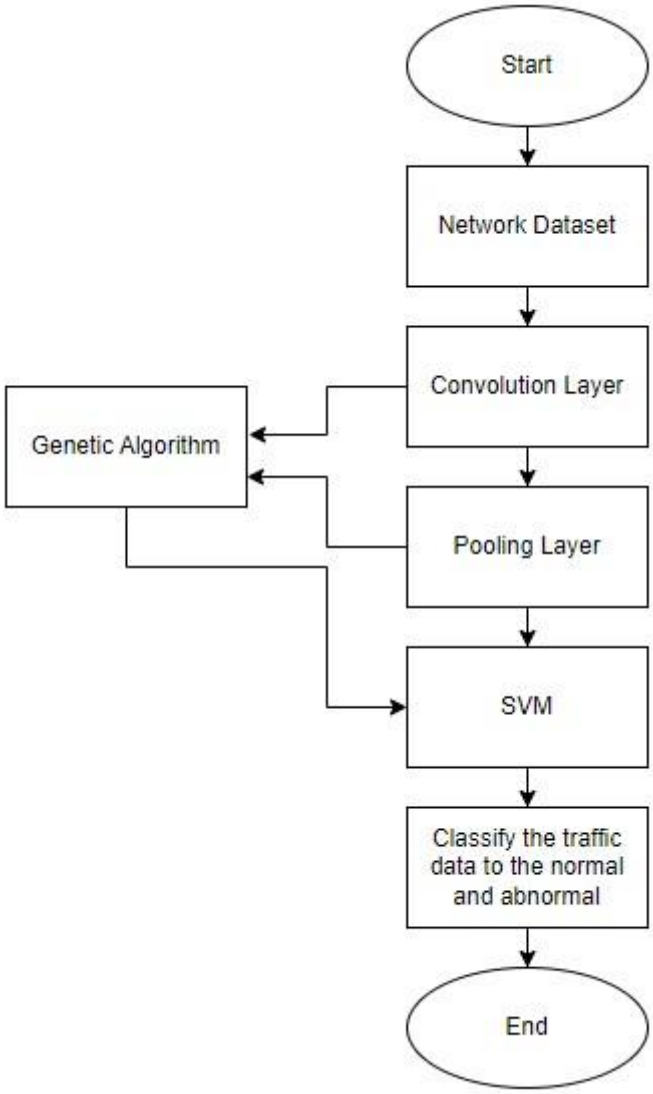

Figure 3.2: Proposed Method.

4. SIMULATION RESULTS

ACCURACY

Table 4.1: Accuracy of SVM with

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	87.84	88.94	88.34
Dataset [53]	85.34	86.23	84.43
Dataset [54]	82.73	83.56	81.46

Table 4.2: Accuracy of KNN

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	84.61	87.33	86.45
Dataset [53]	86.13	83.43	82.56
Dataset [54]	81.34	81.65	80.21

Table 4.3: Accuracy of DT

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	84.61	87.33	86.45
Dataset [53]	86.13	83.43	82.56
Dataset [54]	81.34	81.65	80.21

Table 4.4: Accuracy of RF

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	88.94	89.65	88.43
Dataset [53]	86.45	86.23	83.56
Dataset [54]	84.65	85.65	82.74

Table 4.5: Accuracy of ANN

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	88.94	89.65	88.43
Dataset [53]	86.45	86.23	83.56
Dataset [54]	84.65	85.65	82.74

Table 4.6: Accuracy of AdaBoost

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	88.94	89.65	88.43
Dataset [53]	86.45	86.23	83.56
Dataset [54]	84.65	85.65	82.74

Table 4.7: Accuracy of CNN with Genetic Algorithm

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	99.54	99.65	99.34
Dataset [53]	98.54	97.54	99.85
Dataset [54]	98.54	98.54	97.54

SENSITIVITY

Table 4.8: Sensitivity of SVM

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Cross Validation	Random Subsampling
Dataset [52]	87.84	88.94	88.34
Dataset [53]	85.34	86.23	84.43
Dataset [54]	82.73	83.56	81.46

Table 4.9: Sensitivity of KNN

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	84.61	87.33	86.45
Dataset [53]	86.13	83.43	82.56
Dataset [54]	81.34	81.65	80.21

Table 4.10: Sensitivity of DT

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	84.61	87.33	86.45
Dataset [53]	86.13	83.43	82.56
Dataset [54]	81.34	81.65	80.21

Table 4.11: Sensitivity of RF

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	88.94	89.65	88.43
Dataset [53]	86.45	86.23	83.56
Dataset [54]	84.65	85.65	82.74

Table 4.12: Sensitivity of ANN

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	88.94	89.65	88.43
Dataset [53]	86.45	86.23	83.56
Dataset [54]	84.65	85.65	82.74

Table 4.13: Sensitivity of AdaBoost

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	88.94	89.65	88.43
Dataset [53]	86.45	86.23	83.56
Dataset [54]	84.65	85.65	82.74

Table 4.14: Sensitivity of CNN with Genetic Algorithm

Datasets (%)	Leave-One-Out Cross-Validation	K-fold Validation Cross	Random Subsampling
Dataset [52]	99.54	99.65	99.34
Dataset [53]	98.54	97.54	99.85
Dataset [54]	98.54	98.54	97.54

5. CONCLUSIONS

Cyber security can be defined simply based on the objectives of cyber security. Cyber security; It can be defined as a concept that covers the tools, security guarantees, policies, guidelines, risk management approaches, training and technologies used to protect the properties of institutions, organizations and users against security threats in cyberspace, and the activities within this scope. Cyber security is a new term that has become widespread after 2000 and has a strategic perspective on reducing threats. It is seen as the first security concept in the information revolution. Later, the concept of information security, which is a comprehensive concept, has become widespread. Cybersecurity has grown to include all information processed or stored on computer networks. While the information was dispersed, the problem of information security emerged. Recently, as the interest of the society has shifted to cyber security, there has been a perception that it is equal to internet security. More than that, in fact, the concept of cyber security is the sum of many concepts such as information security, system security and network security. Cyber security can be explained as follows.

Cyber threats; all kinds of cyber-attacks and unauthorized interventions targeting the information assets and equipment of individuals, institutions and countries, disrupting their privacy, security and functioning. Cyber hackers are developing new methods to infiltrate the assets in the cyber environment every day. In the face of these actions carried out in the virtual environment, individuals who are victims of the event are victims both materially and morally. The purpose of cyber-attacks; to cause adverse effects on the principles of confidentiality, integrity and accessibility, which are the three basic elements of ensuring the security of information. The threat potential of cyber terrorism is increasing day by day. Today, terrorist groups benefit from the opportunities offered by technology and are widely used in activities such as propaganda, education, communication and information gathering. The discourse of cyber terrorism started in the early 1990s, when internet technologies began to grow rapidly, the "information society" debate was made, and the studies examining the risks that the USA, which is too dependent on technology and computer network, could face, increased. As a subset of terrorism, cyberterrorism is the use of computers as a weapon, a method, or a target to achieve terrorist goals (Collin, 2004). According to Denning, cyber terrorism is an action or activity aimed at destroying national balance and interests through

the use of electronic tools, computer programs or other electronic communication units in line with information systems.

We used several experiments to classify the network traffic to detect attacks in the network. These experiments combine several deep learning and machine learning techniques. Furthermore, we also applied three validation techniques to evaluate the results. The proposed method presented more than 99% accuracies in several cases.

In future studies, we advises to apply other deep learning techniques such as autoencoders and other optimization algorithms.

REFERENCES

[1] S. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. Chan, “Cost-based modeling for fraud and intrusion detection: results from the jam project, in: DARPA Information Survivability Conference and Exposition,” DISCEX’00, Proc., vol. 2, no. IEEE, 2000, pp. 130–144, 2000.

[2] M. Alkasassbeh, A. B. A. Hassanat, and G. Al-naymat, “Detecting Distributed Denial of Service Attacks Using Data Mining Techniques,” vol. 7, no. 1, pp. 436–445, 2016.

[3] A. Abraham, C. Grosan, and C. Martin-Vide, “Evolutionary design of intrusion detection programs,” Int. J. Netw. Secur., vol. 4, no. 3, pp. 328–339, 2007.

[4] A. M. Brues, “Genetic effects of the atom bomb,” J. Hered., vol. 38, no. 5, pp. 137–137, 1947.

[5] H. Liu, Y. Sun, and M. S. Kim, “Fine-grained DDoS detection scheme based on bidirectional count sketch,” Proc. - Int. Conf. Comput. Commun. Networks, ICCCN, 2011.

[6] C. Khammassi and S. Krichen, “A GA-LR wrapper approach for feature selection in network intrusion detection,” Comput. Secur., vol. 70, pp. 255–277, 2017.

[7] A. Abraham and J. Thomas, “Distributed intrusion detection systems: a computational intelligence approach,” Idea Gr. Inc. Publ. Usa, Chapter, vol. 5, pp. 1–28, 2005.

[8] N. Sharma and S. Mukherjee, “A Novel Multi-Classifier Layered Approach to Improve Minority Attack Detection in IDS,” Procedia Technol., vol. 6, pp. 913–921, 2012.

[9] B. Škrbić and N. Durišić-Mladenović, “Principal component analysis for soil contamination with organochlorine compounds,” Chemosphere, vol. 68, no. 11, pp. 2144–2152, 2007.

[10] N. Patani and R. Patel, “A Mechanism for Prevention of Flooding based DDoS Attack,” vol. 13, no. 1, pp. 101–111, 2017.

[11] L. K. Xr et al., “8VLQJ 6WDFNHG ’ HQRLVLQJ $ XWRHQFRGHU IRU WKH,” pp. 483–488, 2017.

[12] Y. Feng, R. Guo, D. Wang, and B. Zhang, “Research on the active DDoS filtering algorithm based on IP flow,” 5th Int. Conf. Nat. Comput. ICNC 2009, vol. 4, no. October, pp. 628–632, 2009.

[13] A. Meek, “DDoS attacks are getting much more powerful and the Pentagon is scrambling for solutions,” 2015.

[14] J. Mirkovic and P. Reiher, “A taxonomy of DDoS attack and DDoS defense mechanisms,” ACM SIGCOMM Comput. Commun. Rev., vol. 34, no. 2, p. 39, 2004.

[15] S. Taghavi Zargar, J. Joshi, D. Tipper, and S. Member, “A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks,” pp. 1–24, 2013.

[16] T. T. Oo and T. Phyu, “Analysis of DDoS Detection System based on Anomaly Detection System,” 2014.

[17] S. Sinha and M. Sharma, “Simulation and Analysis of DDoS Attacks by Specialized Simulator using Virtualization,” vol. 3, no. 2, pp. 2–4, 2014.

[18] Bojan Kolosnjaji, Apostolis Zarras, George Webster, and Claudia Eckert. Deep learn- ing for classification of malware system call sequences. In Australasian Joint Confer- ence on Artificial Intelligence, pages 137–149. Springer, 2016.

[19] B. Hang, R. Hu, and W. Shi, “An enhanced SYN cookie defence method for TCP DDoS attack,” J. Networks, vol. 6, no. 8, pp. 1206–1213, 2011.

[20] D. Wang, Z. Yufu, and J. Jie, “A multi-core based DDoS detection method,” Proc. - 2010 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol. ICCSIT 2010, vol. 4, pp. 115–118, 2010.

[21] N. Hoque, H. Kashyap, and D. K. Bhattacharyya, “Real-time DDoS attack detection using FPGA,” Comput. Commun., vol. 110, pp. 48–58, 2017.

[22] J. Singh, M. Sachdeva, and K. Kumar, “Detection of DDoS Attacks Using Source IP Based Entropy,” vol. 3, no. 1, pp. 201–210, 2013.

[23] S. M. Lee, D. S. Kim, J. H. Lee, and J. S. Park, “Detection of DDoS attacks using optimized traffic matrix,” Comput. Math. with Appl., vol. 63, no. 2, pp. 501–510, 2012.

[24] H. Rahmani, N. Sahli, and F. Kamoun, “DDoS flooding attack detection scheme based on F-divergence,” Comput. Commun., vol. 35, no. 11, pp. 1380–1391, 2012.

[25] Senirkentli, Güler B., Fatih Ekinci, Erkan Bostanci, Mehmet S. Güzel, Özlem Dağli, Ahmad M. Karim, and Alok Mishra. 2021. "Proton Therapy for Mandibula Plate Phantom" Healthcare 9, no. 2: 167. https://doi.org/10.3390/healthcare9020167.

[26] V. P. Nigam and D. Graupe, “A neural-network-based detection of epilepsy,” Neurol. Res., vol. 26, no. 1, pp. 55–60, 2004.

[27] A. Abraham, U. States, and C. Grosan, “Genetic Systems Programming,” vol. 13, no. May, 2006.

[28] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” Found. Trends® Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014.

[29] K. Ahmed, K. Nia, S. A. Khan, and A. Shaukat, “Identifying Be est Feature Subset for Ca ardiac Arrhy ythmia Classification,” pp. 494–499, 2015.

[30] A. Shenfield, D. Day, and A. Ayesh, “Intelligent intrusion detection systems using artificial neural networks,” ICT Express, vol. 4, no. 2, pp. 95–99, 2018.

[31] A. Kaushik, H. Gupta, and D. S. Latwal, “Impact of Feature Selection and Engineering in the Classification of Handwritten Text,” 2016 Int. Conf. Comput. Sustain. Glob. Dev., pp. 2598–2601, 2016.

[32] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and G. Schneider,
“Generative Recurrent Networks for De Novo Drug Design,” Mol. Inform., vol. 37, no. 1, 2018.

[33] S. Ibrahim, R. Djemal, and A. Alsuwailem, “Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis,” Biocybern. Biomed. Eng., vol. 38, no. 1, pp. 16–26, 2018.

[34] N. Kohli, N. K. Verma, and A. Roy, “SVM based methods for arrhythmia classification in ECG,” 2010 Int. Conf. Comput. Commun. Technol. ICCCT-2010, pp. 486–490, 2010.

[35] V. Sze, Y.-H. Chen, T.-J. Yang, and J. Emer, “Efficient Processing of Deep Neural Networks: A Tutorial and Survey,” vol. 105, no. 12, pp. 2295–2329, 2017.

[36] D. Petkovic et al., “SETAP: Software engineering teamwork assessment and prediction using machine learning,” 2014 IEEE Front. Educ. Conf. Proc., pp. 1–8, 2014.

[37] C. Sobie, C. Freitas, and M. Nicolai, “Simulation-driven machine learning: Bearing fault classification,” Mech. Syst. Signal Process., vol. 99, pp. 403–419, 2018.

[38] J. H. Lee, J. Shin, and M. J. Realff, “Machine learning: Overview of the recent progresses and implications for the process systems engineering field,” Comput. Chem. Eng., 2017.

[39] T. M. Mitchell, (Mcgraw-Hill International Edit) Thomas Mitchell-Machine learningMcGraw Hill Higher Education (1997). .

[40] S. RAY, “Understanding Support Vector Machine algorithm from examples,” 2017. [41] A. M. Karim, Ö. Karal, and F. V Çelebi, “A New Automatic Epilepsy Serious Detection
Method by Using Deep Learning Based on Discrete Wavelet Transform,” no. 4, pp. 15–18, 2018.

[42] Ahmad M. Karim, Mehmet S. Güzel, Mehmet R. Tolun, Hilal Kaya, Fatih V. Çelebi, A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing, Biocybernetics and Biomedical Engineering, Volume 39, Issue 1,2019,Pages 148-159, ISSN 0208-5216, https://doi.org/10.1016/j.bbe.2018.11.004.

[43] S. M. Hosseini Bamakan, H. Wang, and Y. Shi, “Ramp loss K-Support Vector Classification-Regression; a robust and sparse multi-class approach to the intrusion detection problem,” Knowledge-Based Syst., vol. 126, pp. 113–126, 2017.

[44] A. M. Karim, F. V. Çelebi, and A. S. Mohammed, “Software Development for Blood
Disease Expert System,” Lecture Notes on Empirical Software Engineering, vol. 4, no. 3, pp. 179–183, 2016.

[45] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V Çelebi, “A New Generalized Deep Learning Framework Combining Sparse Auto-encoder and Taguchi Method for Novel Data Classification and Processing,” pp. 1–22.

[46] L. Wang, C. Wang, W. Du et al., “Parameter optimization of a four-legged robot to improve motion trajectory accuracy using signal-to-noise ratio theory,” Robotics and Computer-Integrated Manufacturing, vol. 51, pp. 85–96, 2018..

[47] Ahmad Karim, Development of secure Internet of Vehicle Things (IoVT) for smart transportation system, Computers and Electrical Engineering, Volume 102, 2022, 108101, ISSN 0045-7906.

[48] Alex Sherstinsky, Fundamentals of Recurrent Neural Network (RNN) and Long ShortTerm Memory (LSTM) network, Physica D: Nonlinear Phenomena, Volume 404, 2020, 132306, ISSN 0167-2789.

[49] Karim AM, Kaya H, Alcan V, Sen B, Hadimlioglu IA. New Optimized Deep Learning
Application for COVID-19 Detection in Chest X-ray Images. *Symmetry*. 2022; 14(5):1003.

[50] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, Yinhai Wang, Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values, Transportation Research Part C: Emerging Technologies, Volume 118, 2020, 102674, ISSN 0968-090X.

[51] Ying Chen, Voltages prediction algorithm based on LSTM recurrent neural network, Optik, Volume 220, 2020, 164869, ISSN 0030-4026.

[52] Neculai Andrei, A Dai–Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization, Applied Mathematics Letters, Volume 21, Issue 2, 2008, Pages 165-171, ISSN 0893-9659.

[53] Muhammed Maruf Öztürk, İbrahim Arda Cankaya, Deniz İpekçi, Optimizing echo state network through a novel fisher maximization based stochastic gradient descent, Neurocomputing, Volume 415, 2020, Pages 215-224, ISSN 0925-2312.

[54] Karim, A.M., Mishra, A. (2022). Novel COVID-19 Recognition Framework Based on Conic Functions Classifier. In: Garg, L., Chakraborty, C., Mahmoudi, S., Sohmen, V.S. (eds) Healthcare Informatics for Fighting COVID-19 and Future Epidemics. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-72752-9_1.